Occam's Razor, Boltzmann's Brain and Wigner's Friend

Charles H. Bennett* *(IBM Research)*
including joint work with
C. Jess Riedel* *(Perimeter Institute)*
and old joint work with
Geoff Grinstein *(IBM, retired)*

Public Lecture, Niels Bohr Institute
24 October 2019

*Supported in part by a grant from the
John Templeton Foundation

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.

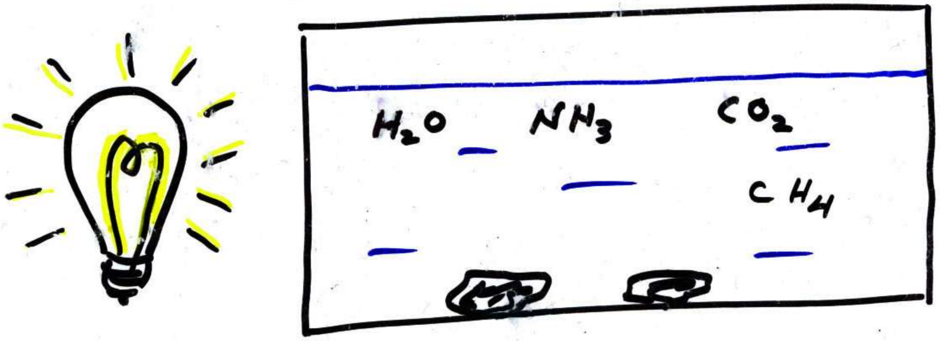
- The relation between *Dynamics*—the spontaneous motion or change of a system obeying physical laws—and *Computation*—a programmed sequence of mathematical operations

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

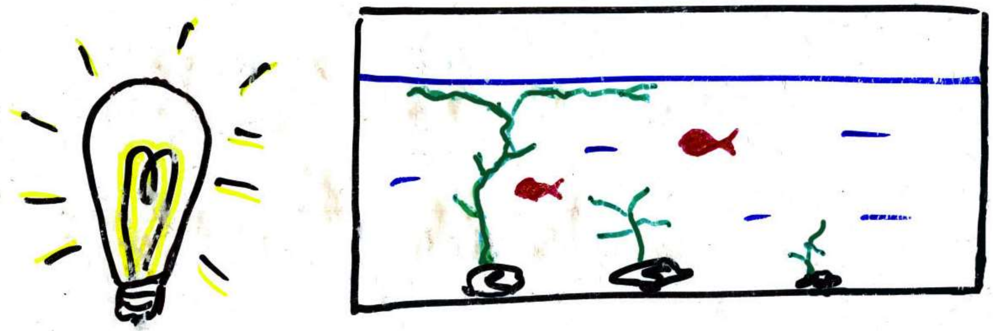
- True and False evidence—the Boltzmann Brain problem at equilibrium and in modern cosmology

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

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



Much later



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

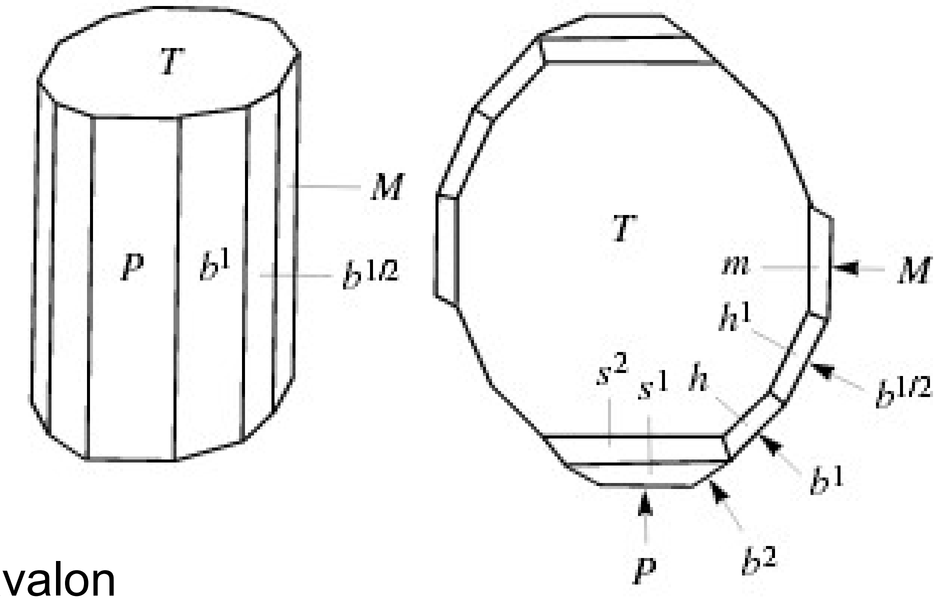
A cartoon by Sidney Harris shows a group of cosmologists pondering an apparent typicality violation

"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."

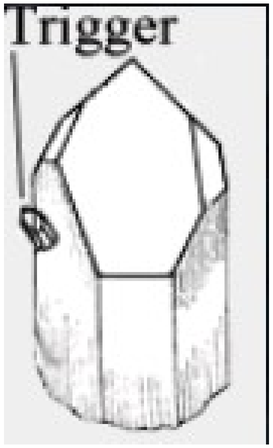
(This cartoon is not too far from problems that actually come up in current cosmology)

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.



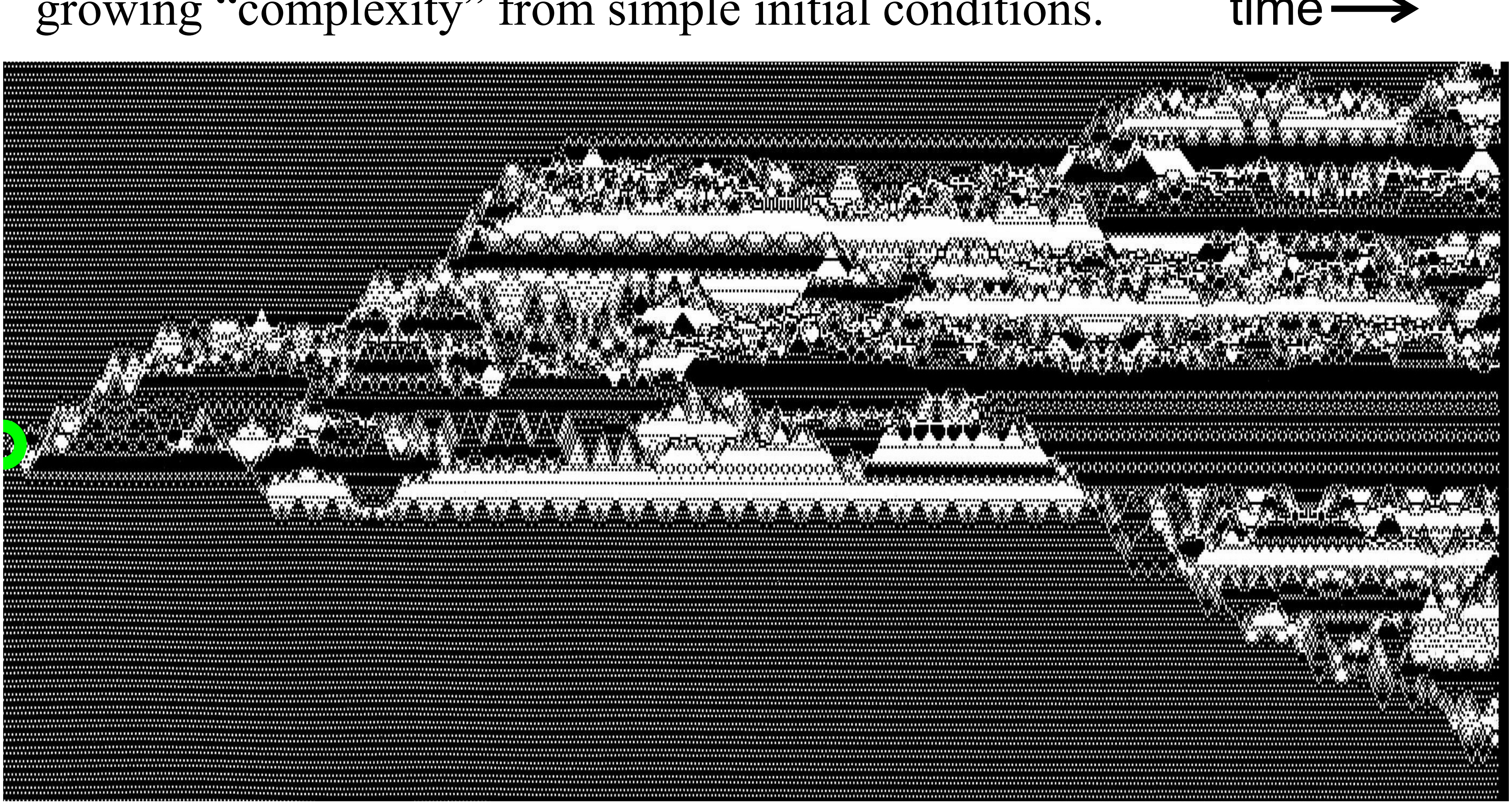
http://www.neatstuff.net/avalon /texts/Quartz-Configurations.html



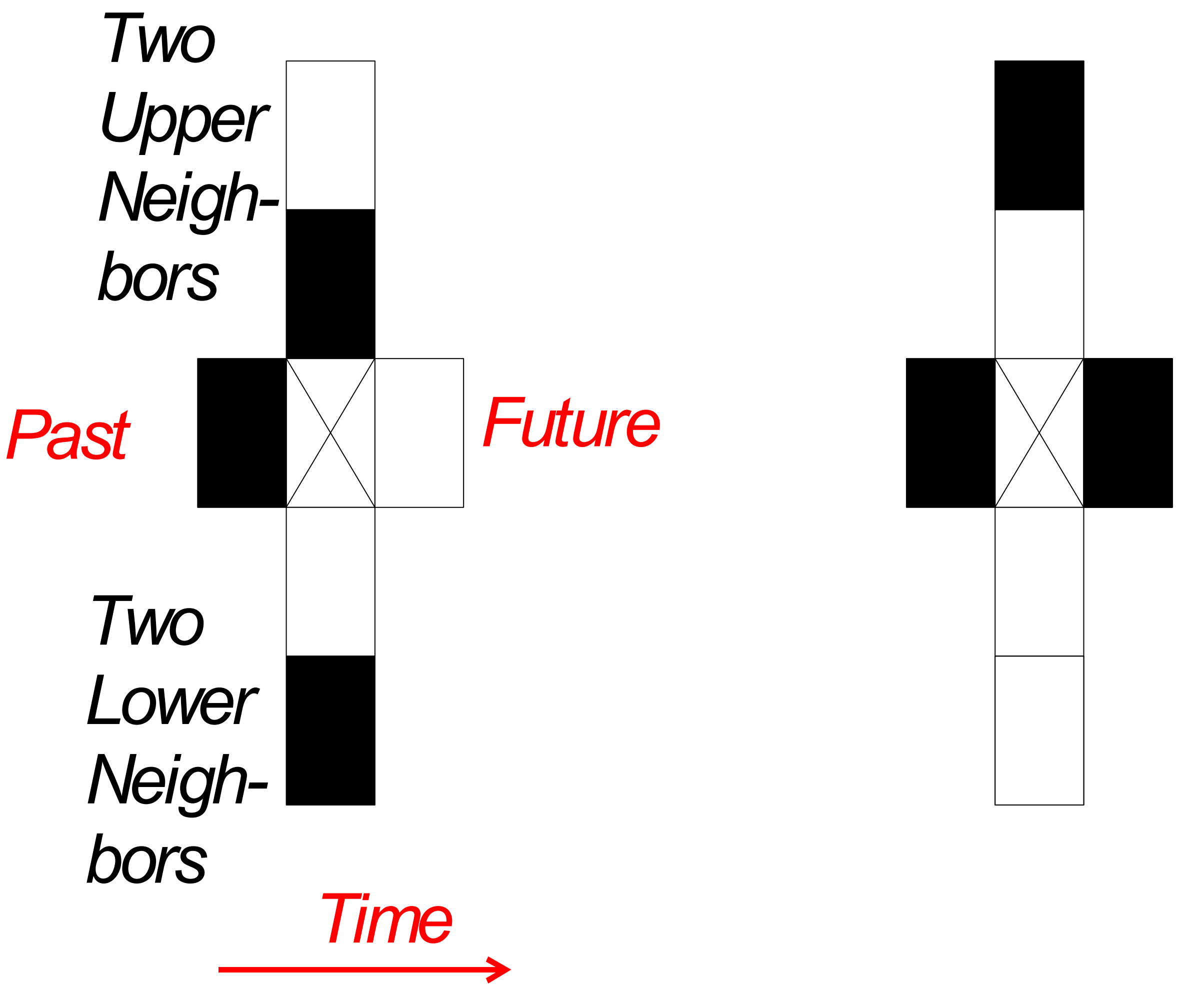
TRIGGER CRYSTALS:
have a **smaller crystal growing out from them**. This `trigger' can be gently squeezed to activate the power of the crystal and strengthen its attributes. These are just used for a surge of a particular kind of energy.

To understand molecules, learn to think like one.

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

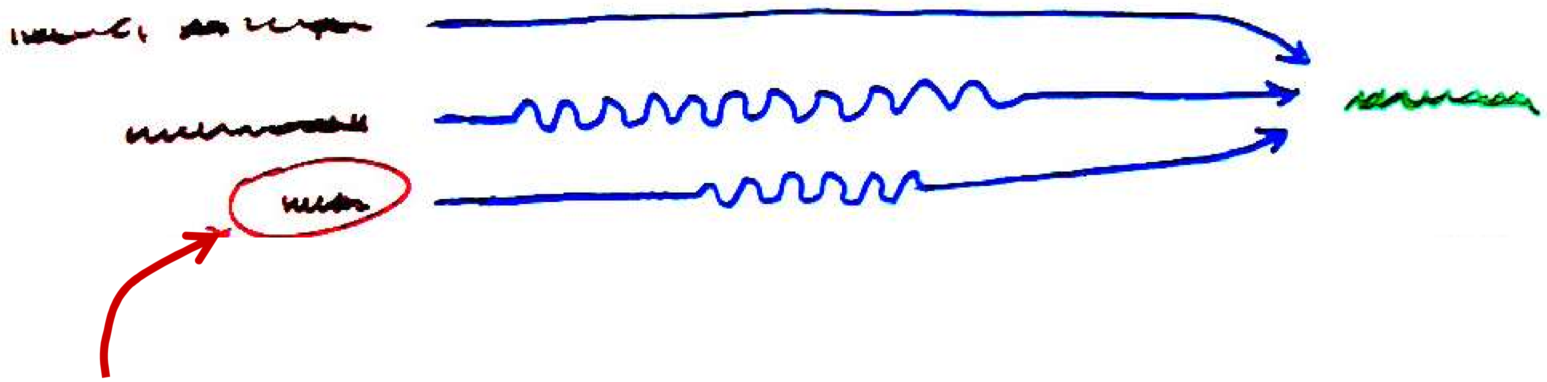


Small irregularity (green) in otherwise periodic initial condition produces a complex deterministic wake.



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

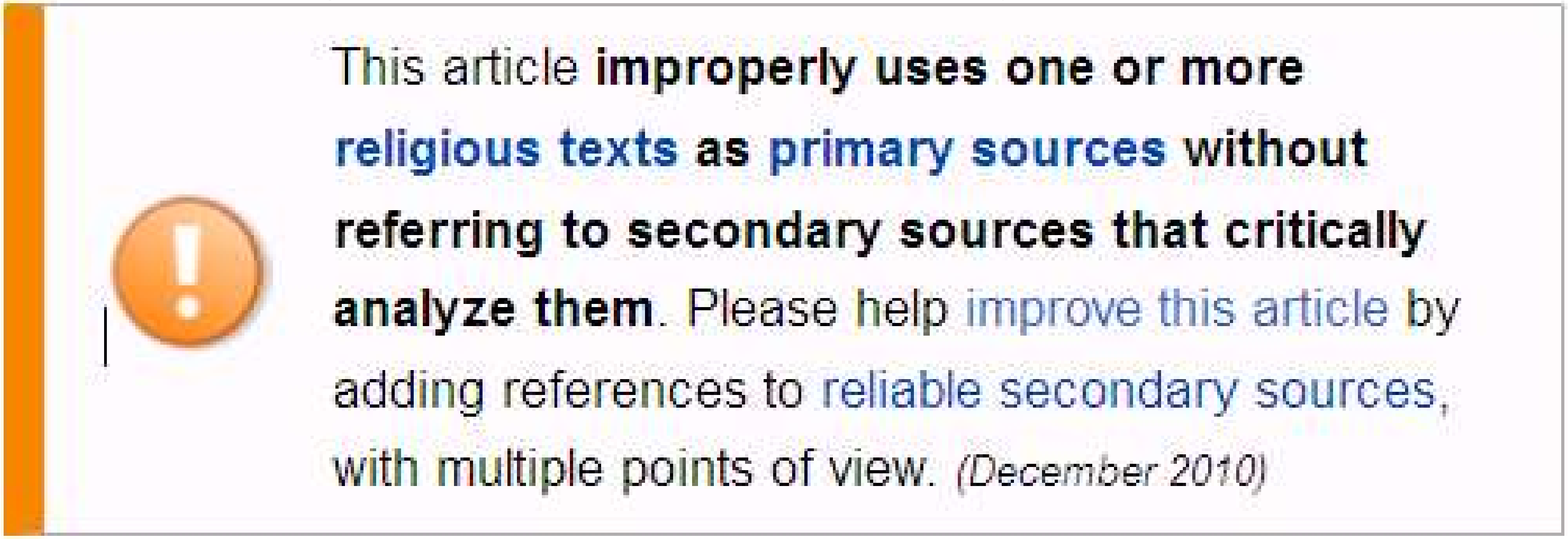
The most economical hypothesis is to be preferred, even if the deductive path connecting it to the phenomena it explains is long and complicated.

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

Original form of Occam's Razor:

> "For nothing ought to be posited without a reason given, unless it is self-evident, or known by experience, or proved by the authority of Sacred Scripture"
>
> *William of Ockham (ca.1287 – 1347)*

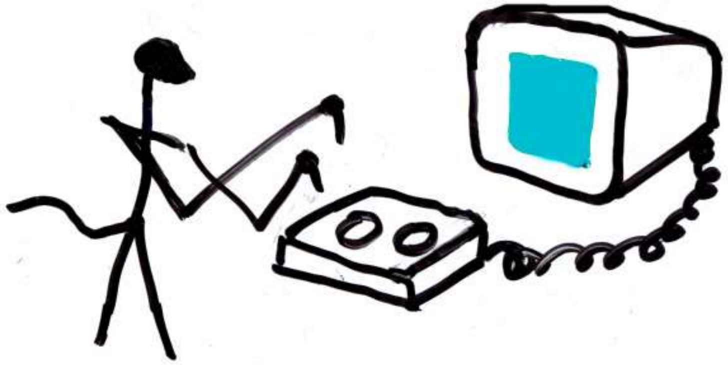
Scriptures get less respect nowadays



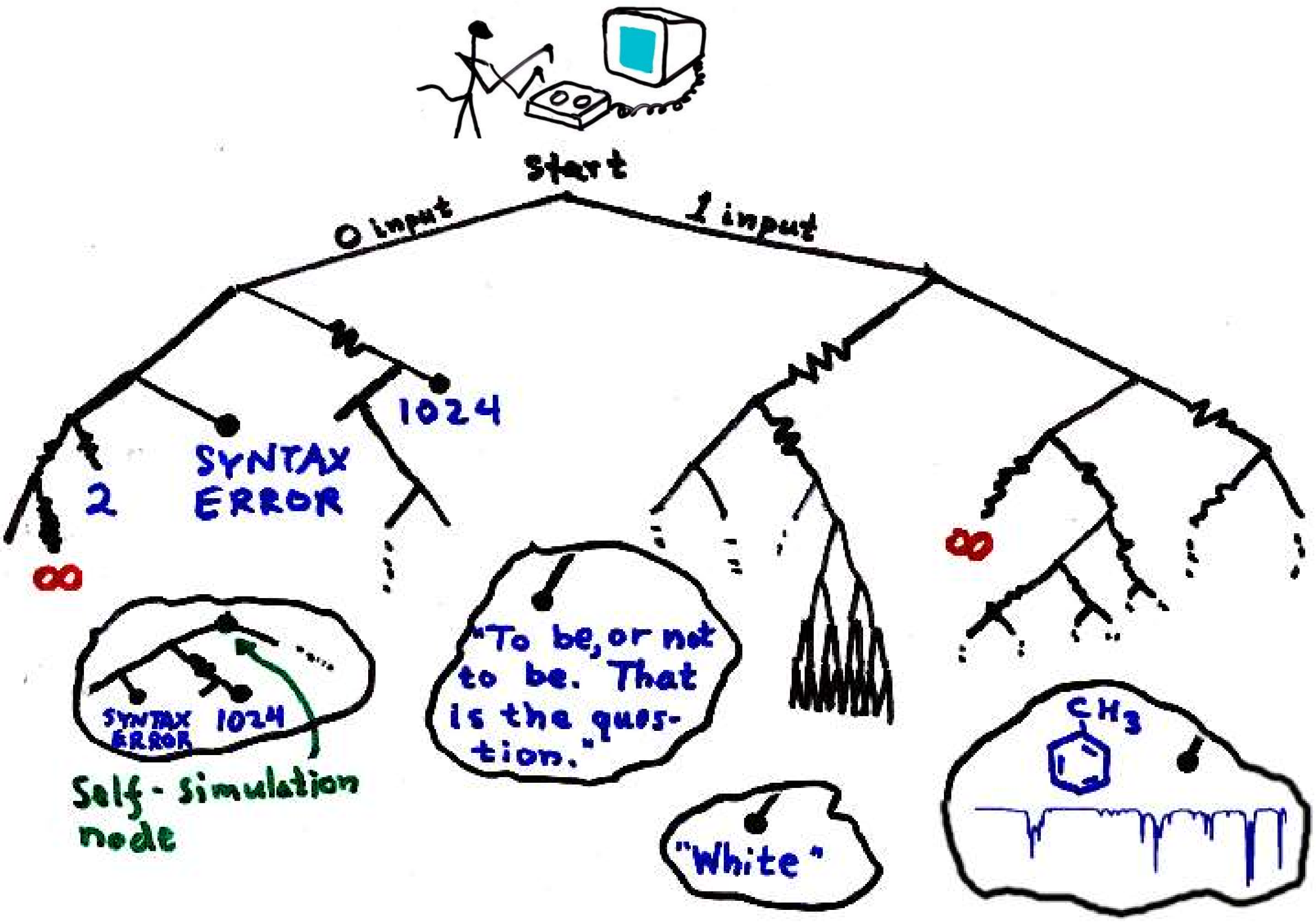
This article **improperly uses one or more religious texts as primary sources without referring to secondary sources that critically analyze them**. Please help improve this article by adding references to reliable secondary sources, with multiple points of view. *(December 2010)*

(Wikipedia warning on early version of Mormon Cosmology article)

Algorithmic information uses a computerized version of the old idea of a monkey at a typewriter eventually typing the works of Shakespeare.

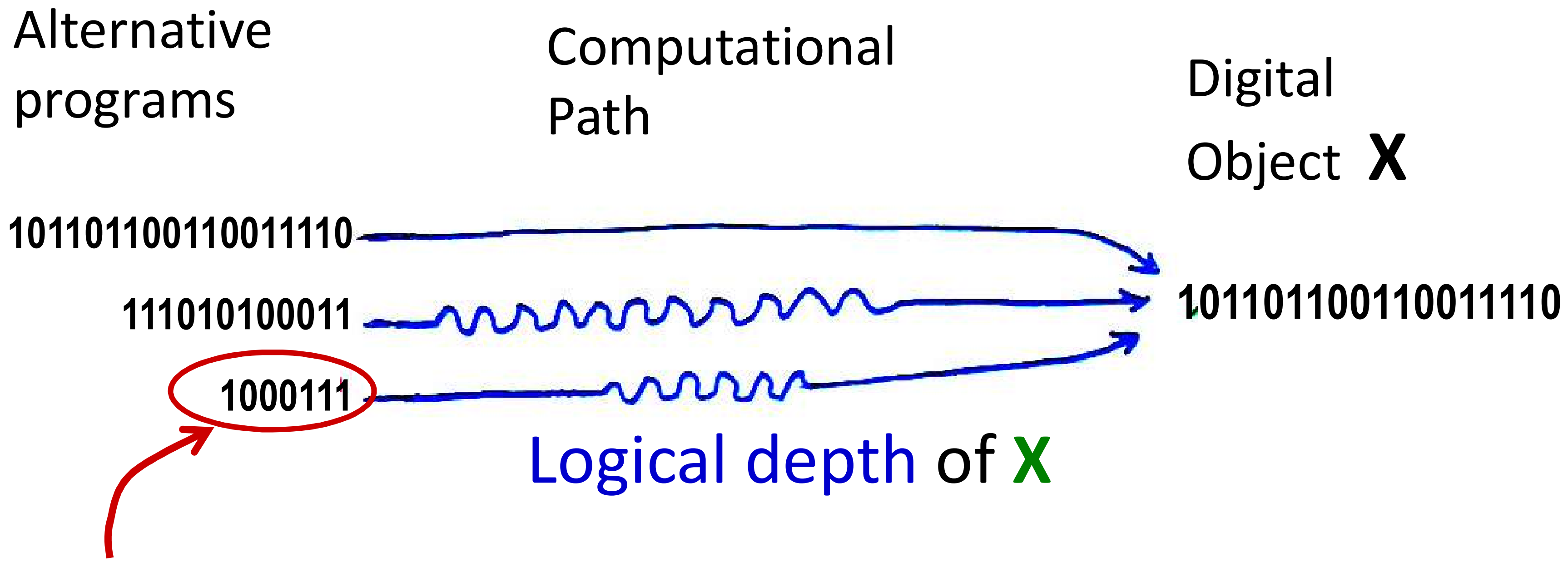


A monkey randomly typing 0s and 1s into a universal binary computer has some chance of getting it to do any computation, produce any output.



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**.



Alternative programs

Computational Path

Digital Object **X**

101101100110011110

111010100011

1000111

101101100110011110

Logical depth of **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.

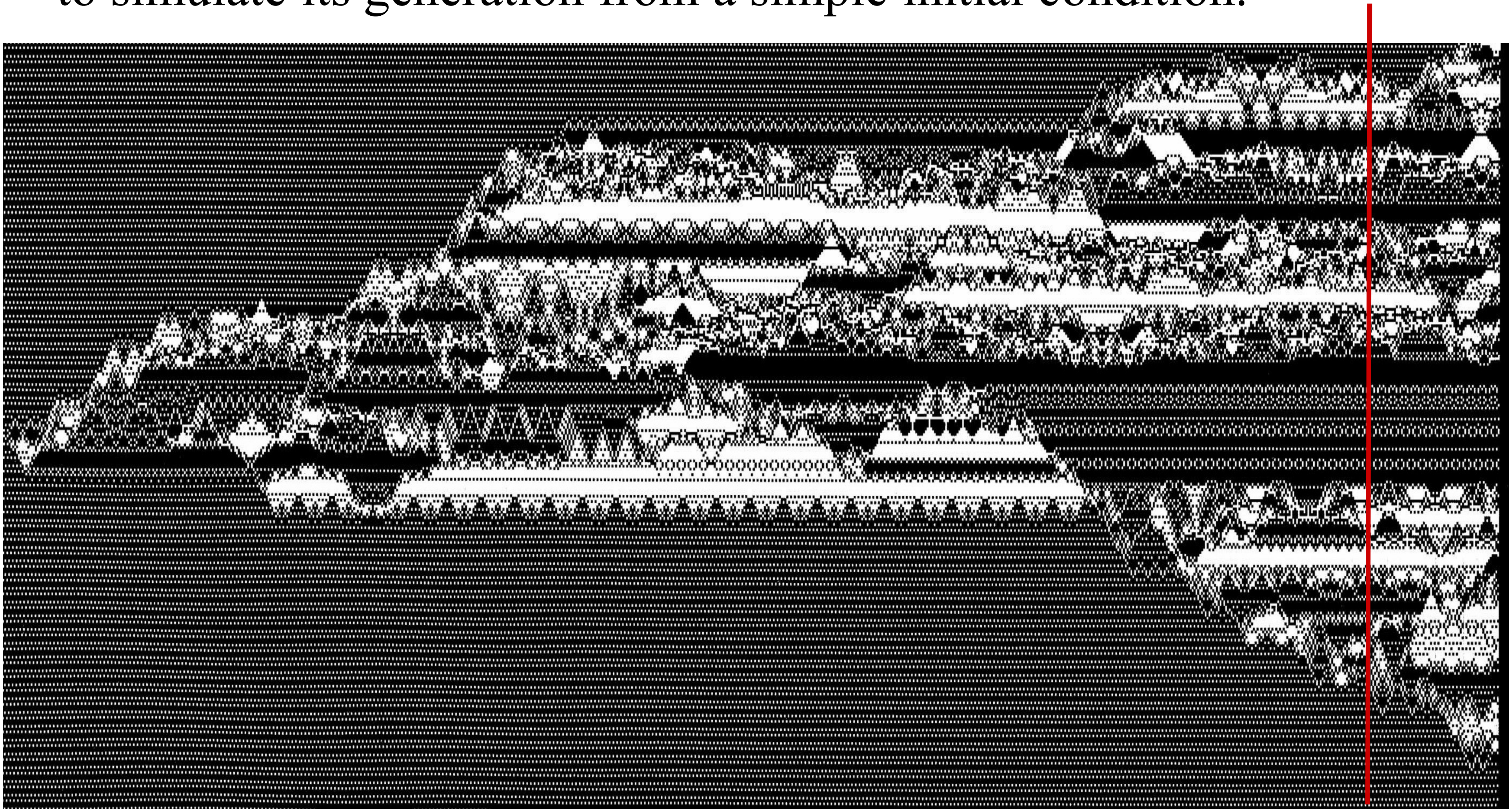
A trivially orderly sequence like 111111… is logically shallow because it can be computed rapidly from a short description.

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.

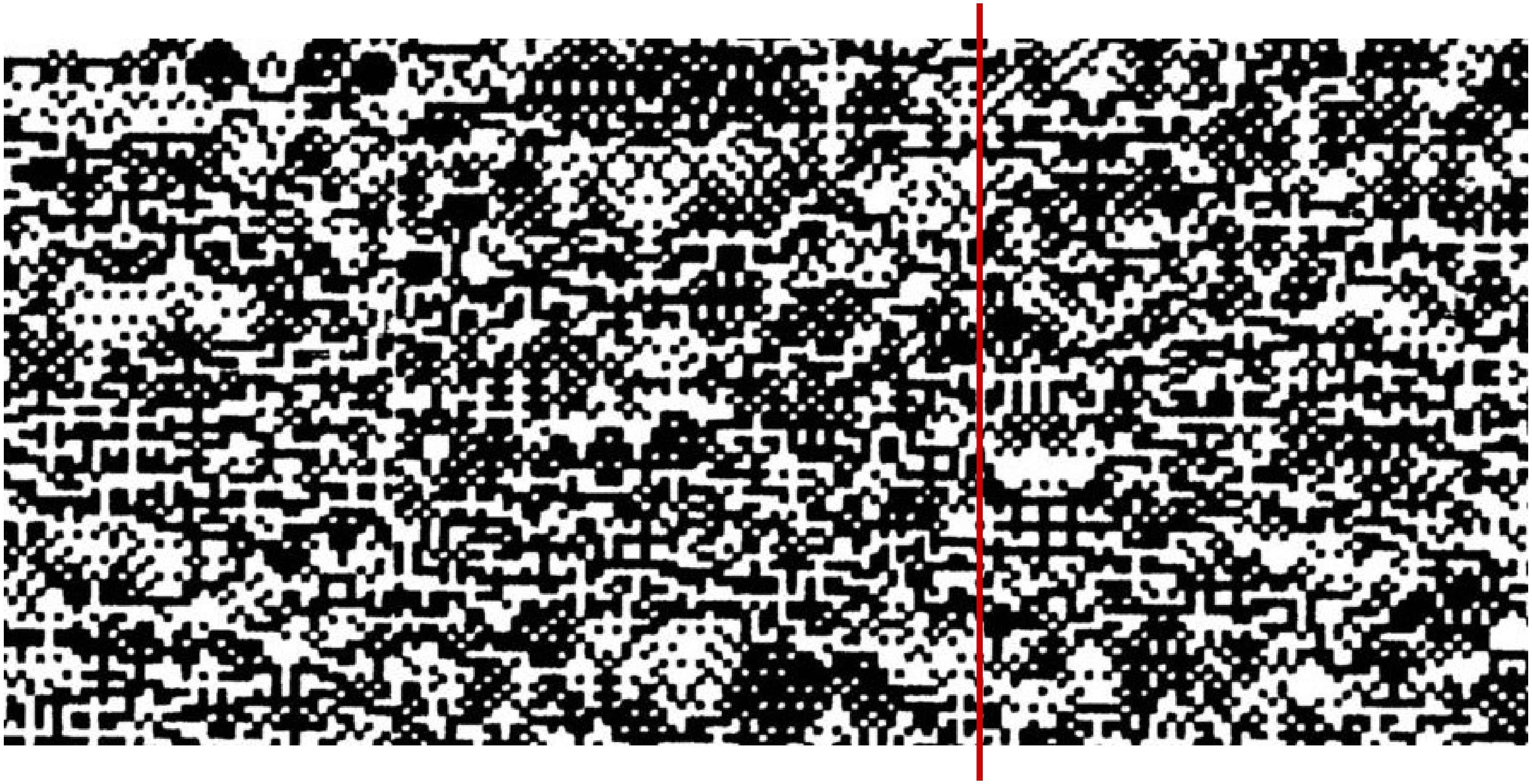
Trivial semi-orderly sequences, such as an alternating sequence of 0's and random bits, are also shallow, since they are rapidly computable from their random part.

(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?

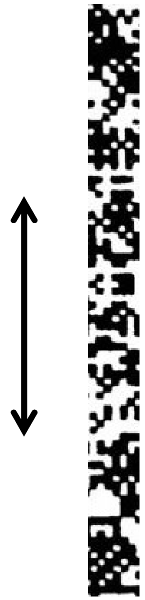
Because to specify the state via a simulation of its actual history would involve naming the exact **number** of steps to run the simulation.

This number is typically very large, requiring about $\boldsymbol{n}$ bits to describe.

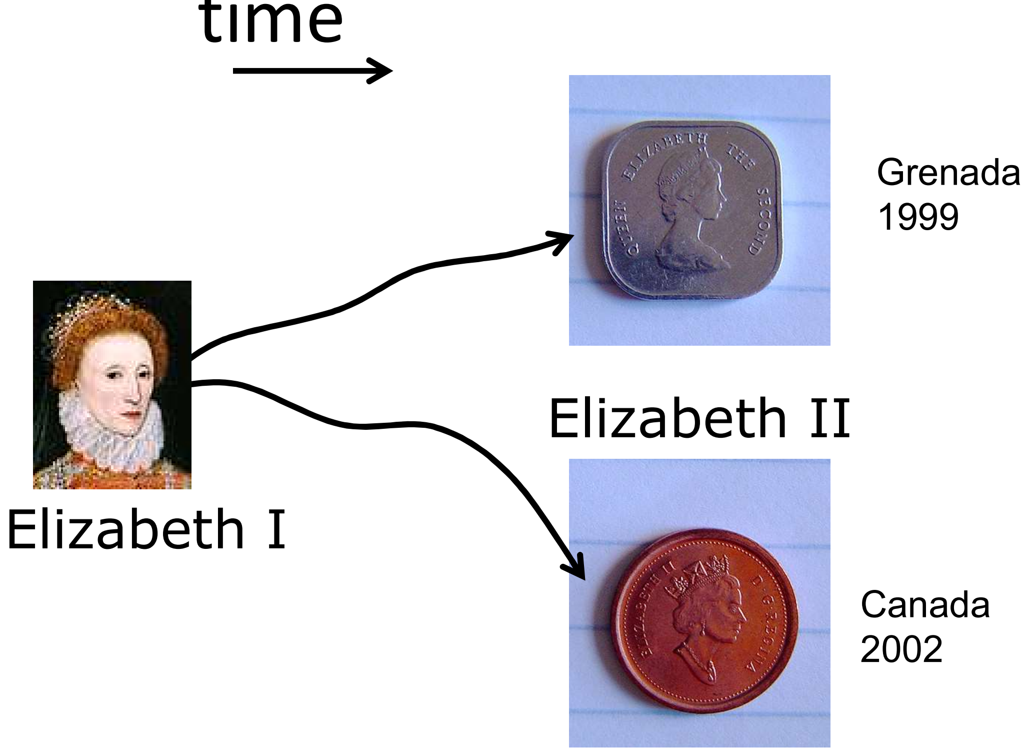
Therefore the actual history is no more plausible (in terms of Occam’s razor) than a “print program” that simply outputs the state from a verbatim description.

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

Correlations mediated through present only

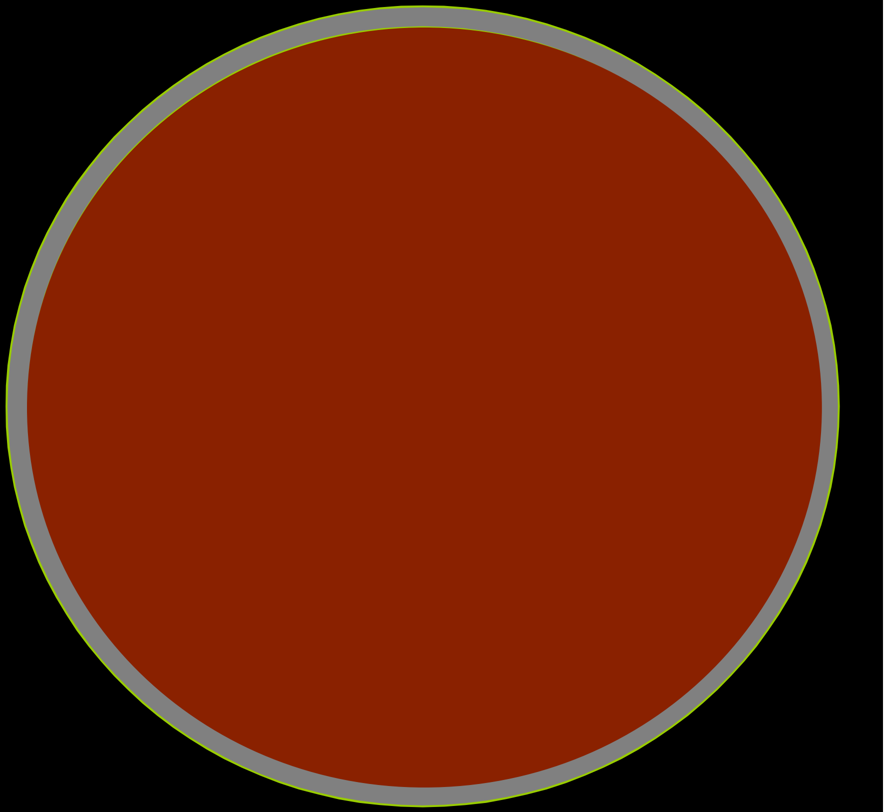


By contrast, in a non-equilibrium world, local dynamics can generically give rise to long range correlations, mediated through a V-shaped path in space-time representing a common history.



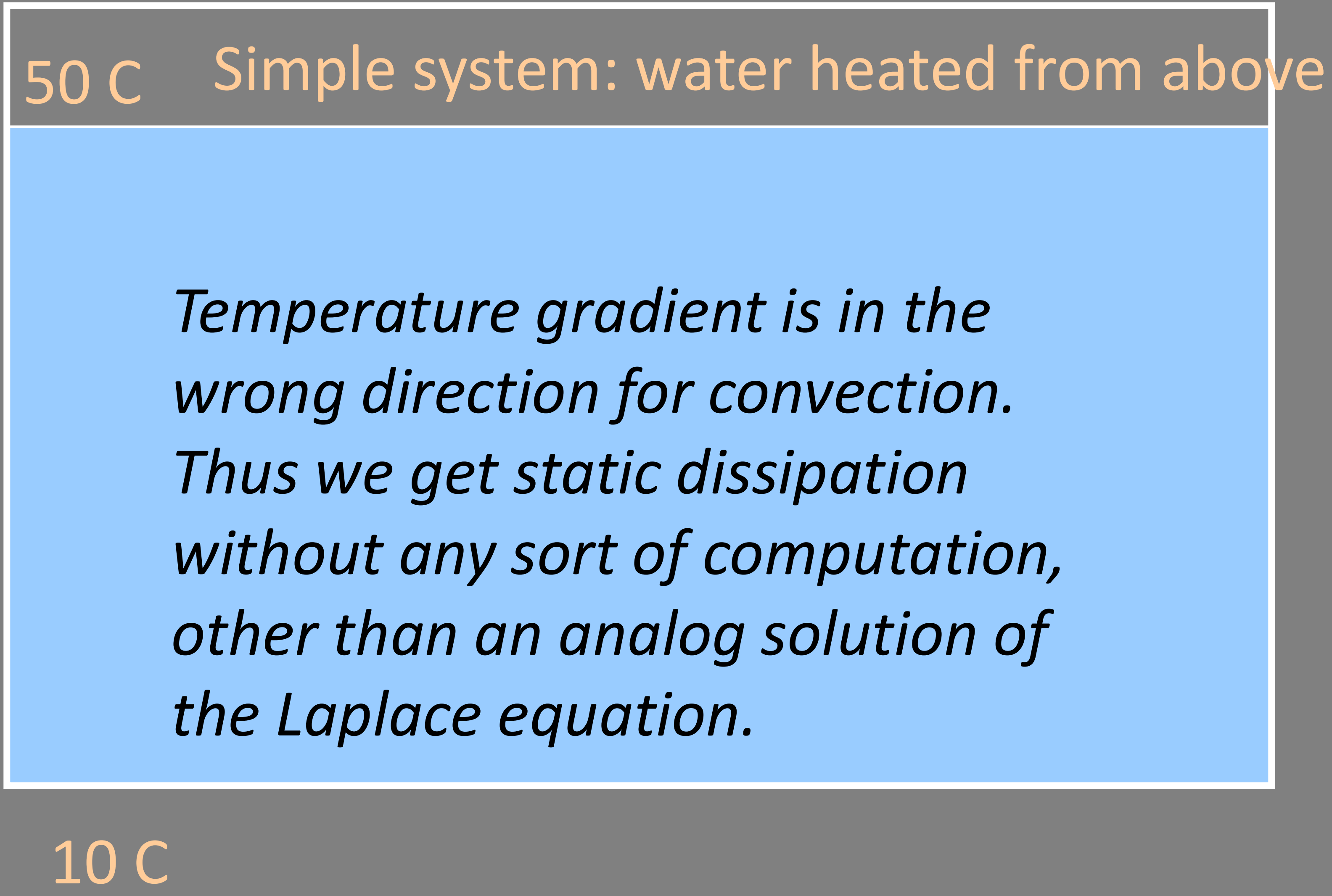
The cellular automaton is a classical toy model, but real systems with fully quantum dynamics behave similarly, losing their complexity, their long-range correlations and even their classical phenomenology as they approach equilibrium.

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.

Dissipation without Computation



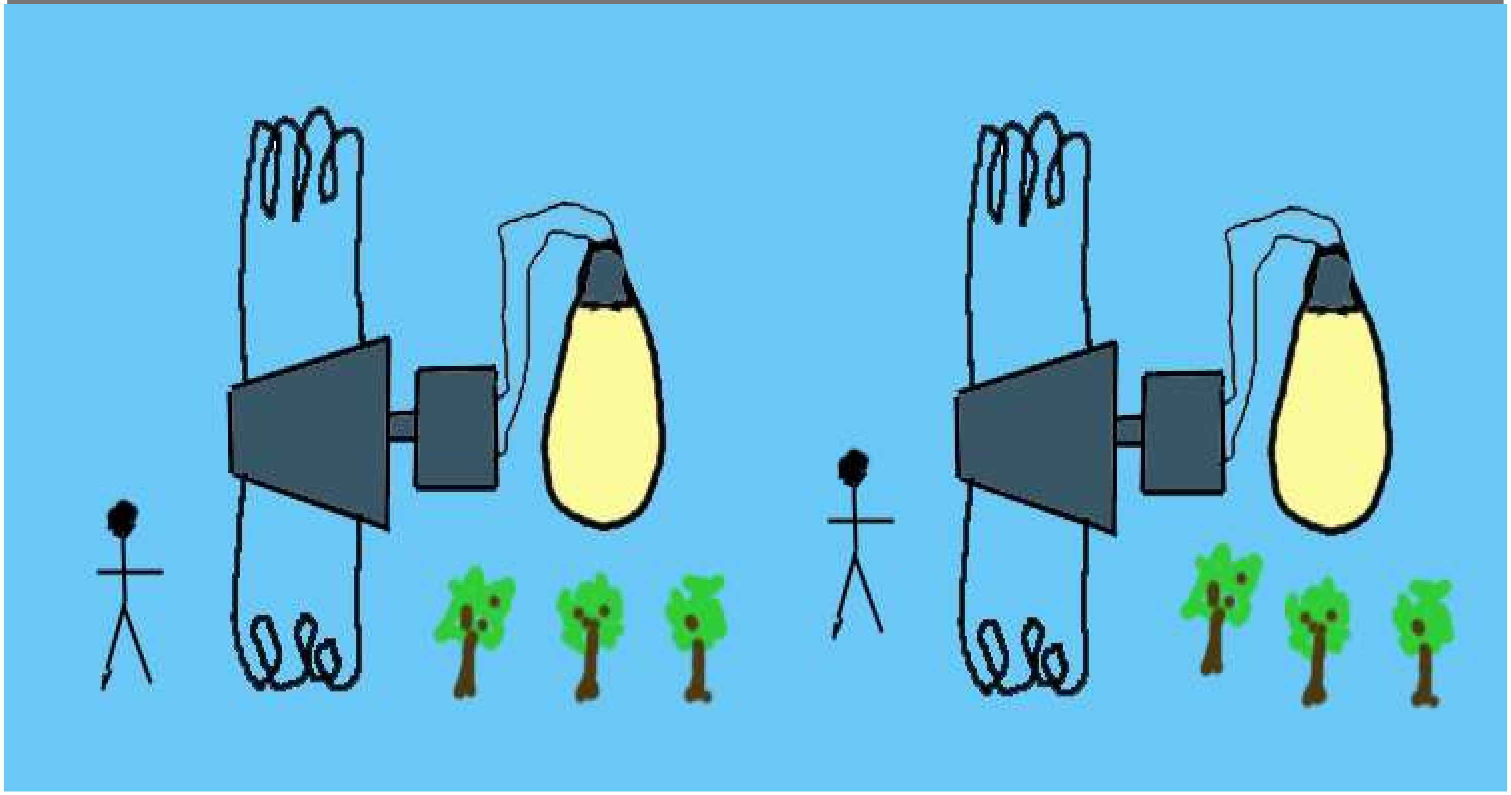
50 C

Simple system: water heated from above

Temperature gradient is in the wrong direction for convection. Thus we get static dissipation without any sort of computation, other than an analog solution of the Laplace equation.

10 C

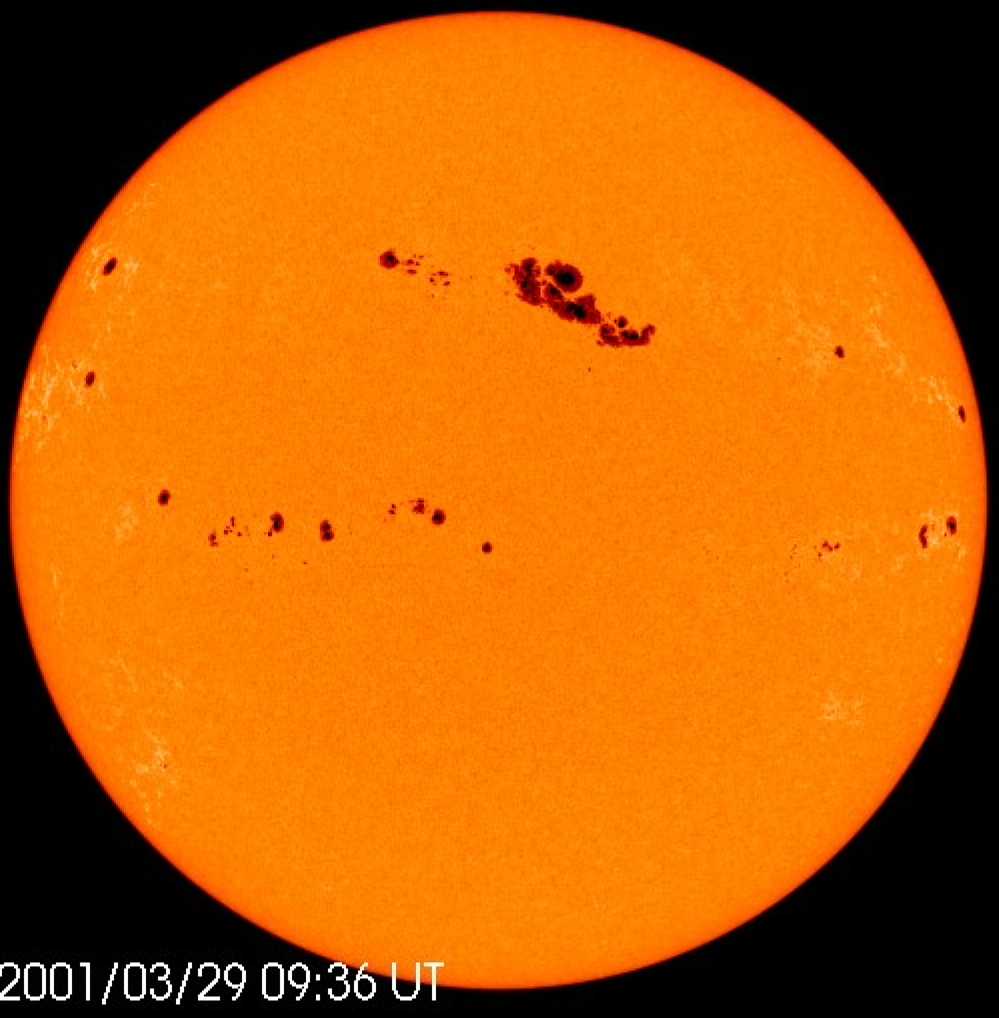
50 C But if the water has impurities



10 C Turbine civilization can maintain and repair itself, do universal computation.

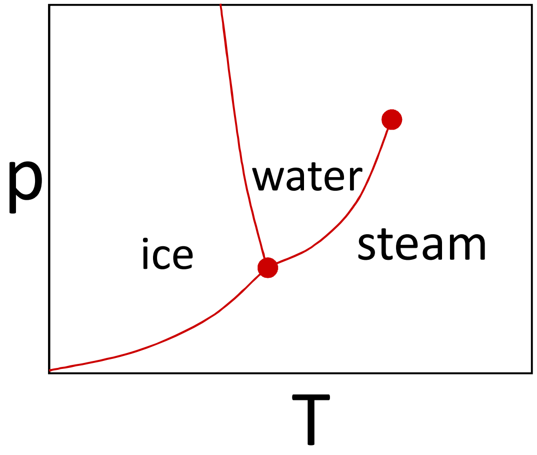
Are some dissipative environments so hot, so rapidly mixing, as to preclude long term memory? Hard to say.

Are some dissipative environments capable of supporting long term memory, but not depth?



Biologically, are there environments where complexity confers no selective advantage and which therefore support only simple life, without niche ecology or the opportunity for preadaptation?

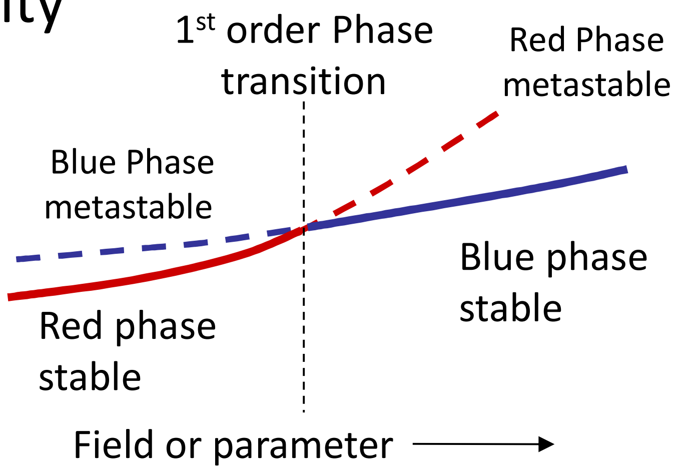
How strong is the connection between disequilibrium and complexity, in the sense of logical depth?



Are thermal equilibrium states generically shallow? Classically 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.

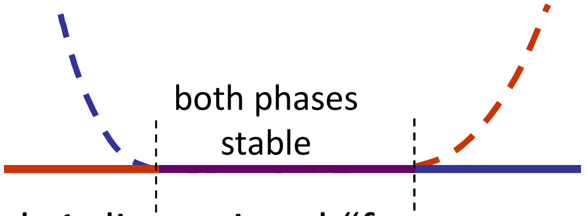
=> No long term memory

=> Equilibrium depth remains bounded in large N limit



Quantum Exceptions? E.g. toric code in 3d or more gives long term memory but not complex equilibrium states.

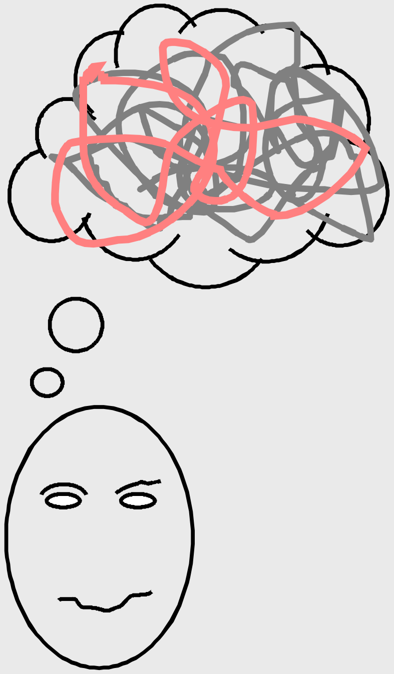
But dissipative systems are truly exempt from the Gibbs phase rule (BG ‘85)



d+1 dimensional “free energy” from system’s transfer matrix

Ordinary classical information, such as one finds in a book, can be copied at will and is not disturbed by reading it.

Quantum information is more like the information in a dream



- Trying to describe your dream changes your memory of it, so eventually you forget the dream and remember only what you've said about it.

- You cannot prove to someone else what you dreamed.

- You can lie about your dream and not get caught.

But unlike dreams, quantum information obeys well-known laws.

Despite the differences there are important similarities between classical and quantum information

All (classical) information is reducible to bits 0 and 1.

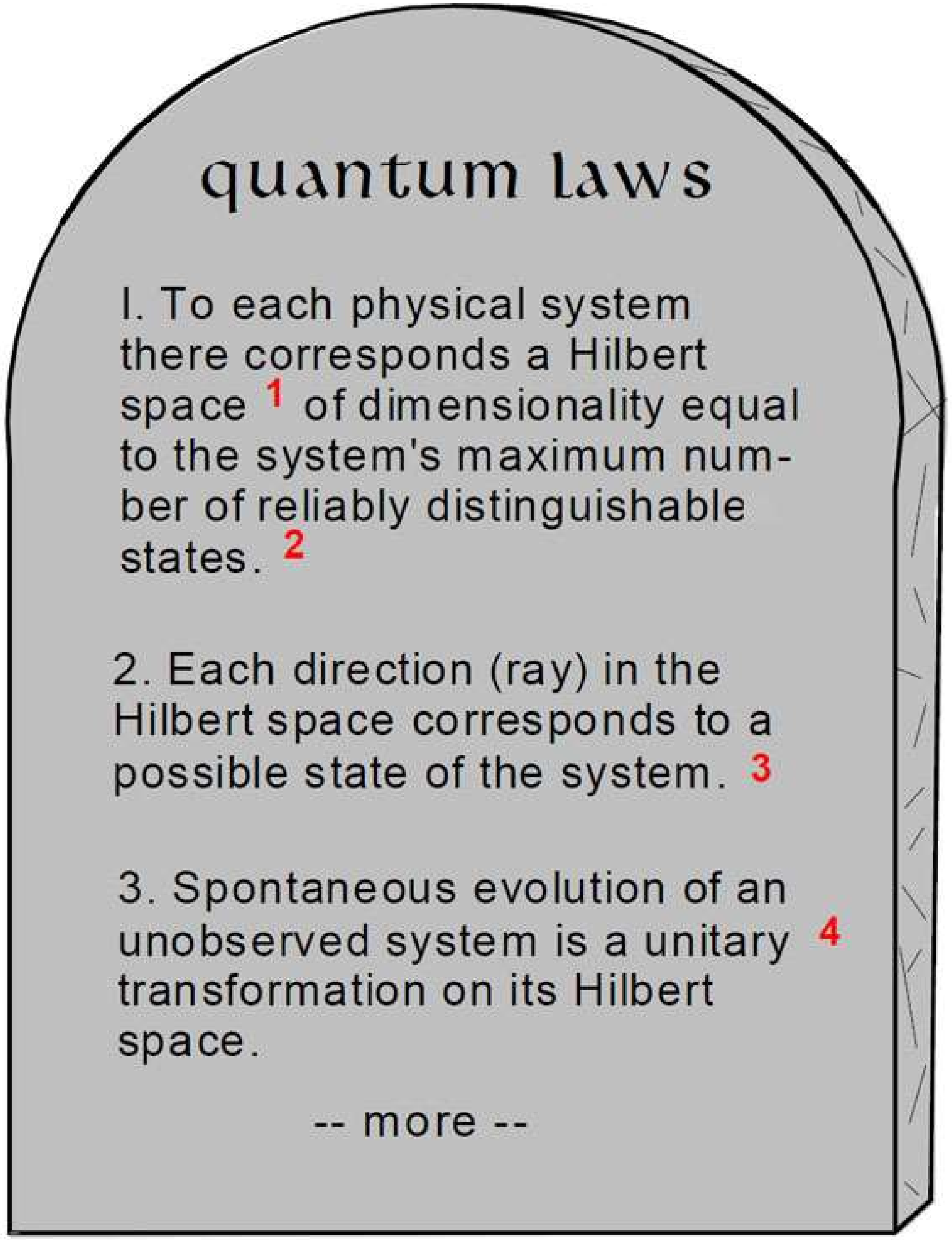
All processing of it can be done by simple logic gates (NOT, AND) acting on bits one and two at a time.

Bits and gates are fungible (independent of physical embodiment), making possible Moore’s law.

Quantum information is reducible to **qubits** i.e. two-state quantum systems such as a photon's polarization or a spin-1/2 atom.

Quantum information processing is reducible to one- and two-qubit gate operations.

Qubits and quantum gates are fungible among different quantum systems



quantum laws

I. To each physical system there corresponds a Hilbert space [1] of dimensionality equal to the system's maximum number of reliably distinguishable states. [2]

2. Each direction (ray) in the Hilbert space corresponds to a possible state of the system. [3]

3. Spontaneous evolution of an unobserved system is a unitary [4] transformation on its Hilbert space.

-- more --

1. A linear vector space with complex coefficients and inner product $< \phi \mid \psi > \; = \Sigma \, \phi^*_i \, \psi_i$

2. For polarized photons two, e.g. vertical and horizonal

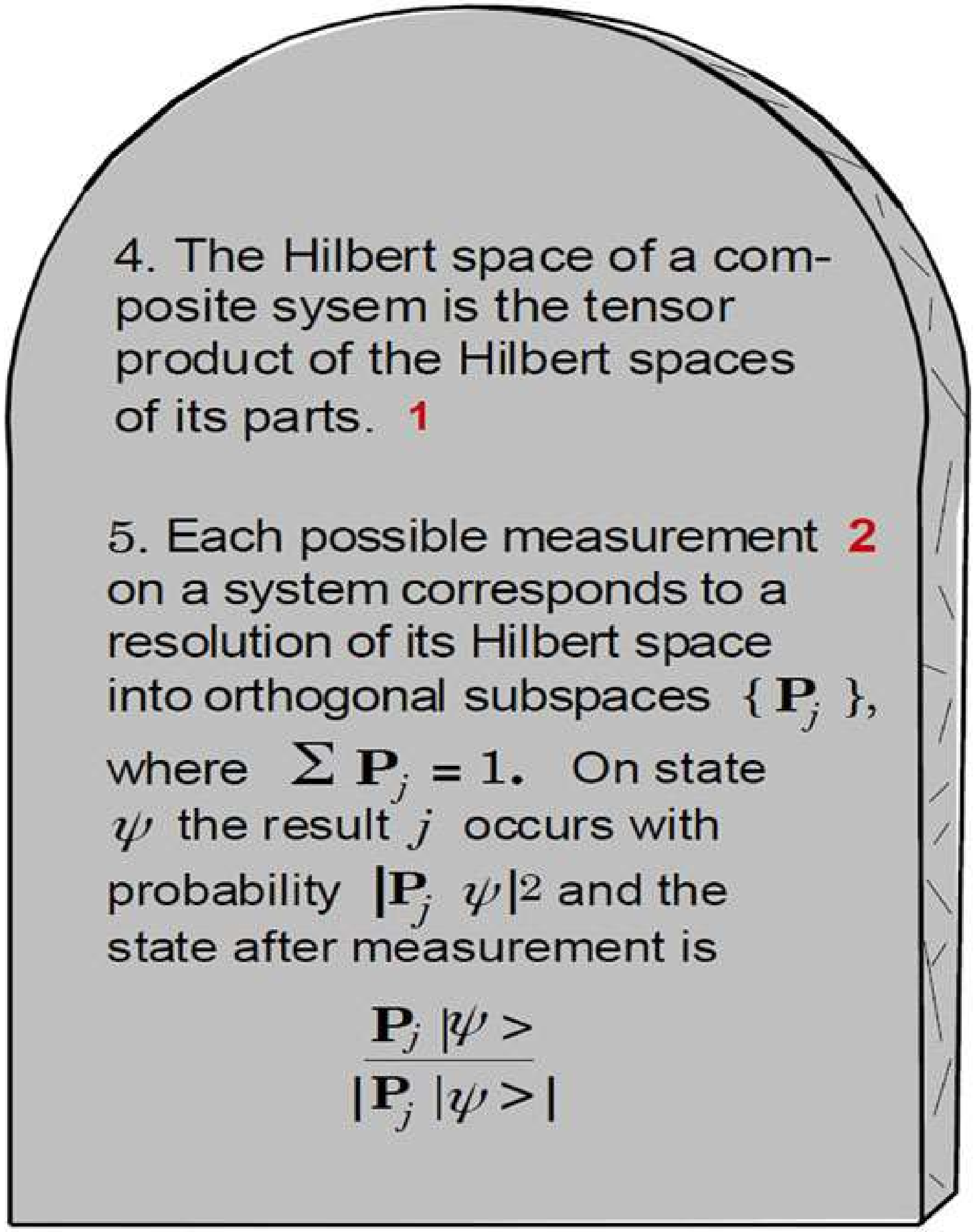
↔ $= \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ↕ $= \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

3. E.g. for photons, other polarizations

⤢ $= \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ⤡ $= \begin{pmatrix} +1 \\ -1 \end{pmatrix}$

↻ $= \begin{pmatrix} i \\ 1 \end{pmatrix}$ ↺ $= \begin{pmatrix} i \\ -1 \end{pmatrix}$

4. Unitary = Linear and inner-product preserving.



4. The Hilbert space of a composite sysem is the tensor product of the Hilbert spaces of its parts. [1]

5. Each possible measurement [2] on a system corresponds to a resolution of its Hilbert space into orthogonal subspaces $\{\mathbf{P}_j\}$, where $\Sigma\,\mathbf{P}_j = 1.$ On state ψ the result j occurs with probability $|\mathbf{P}_j\ \psi|^2$ and the state after measurement is

$$\frac{\mathbf{P}_j\ |\psi>}{|\mathbf{P}_j\ |\psi>|}$$

[1]. Thus a two-photon system can exist in "product states" such as $\leftrightarrow\leftrightarrow$ and $\leftrightarrow$ ⤢ but also in "entangled" states such as

$$\frac{\leftrightarrow\leftrightarrow\ \ -\ \updownarrow\updownarrow}{\sqrt{2}}$$

in which neither photon has a definite state even though the pair together does

[2] Believers in the "many worlds interpretation" reject this axiom as ugly and unnecessary. For them measurement is just a unitary evolution producing an entangled state of the system and measuring apparatus. For others, measurement causes the system to behave probabilistically and forget its pre-measurement state, unless that state happens to lie entirely within one of the subspaces $\mathbf{P}_j$.

The central principle of quantum mechanics is

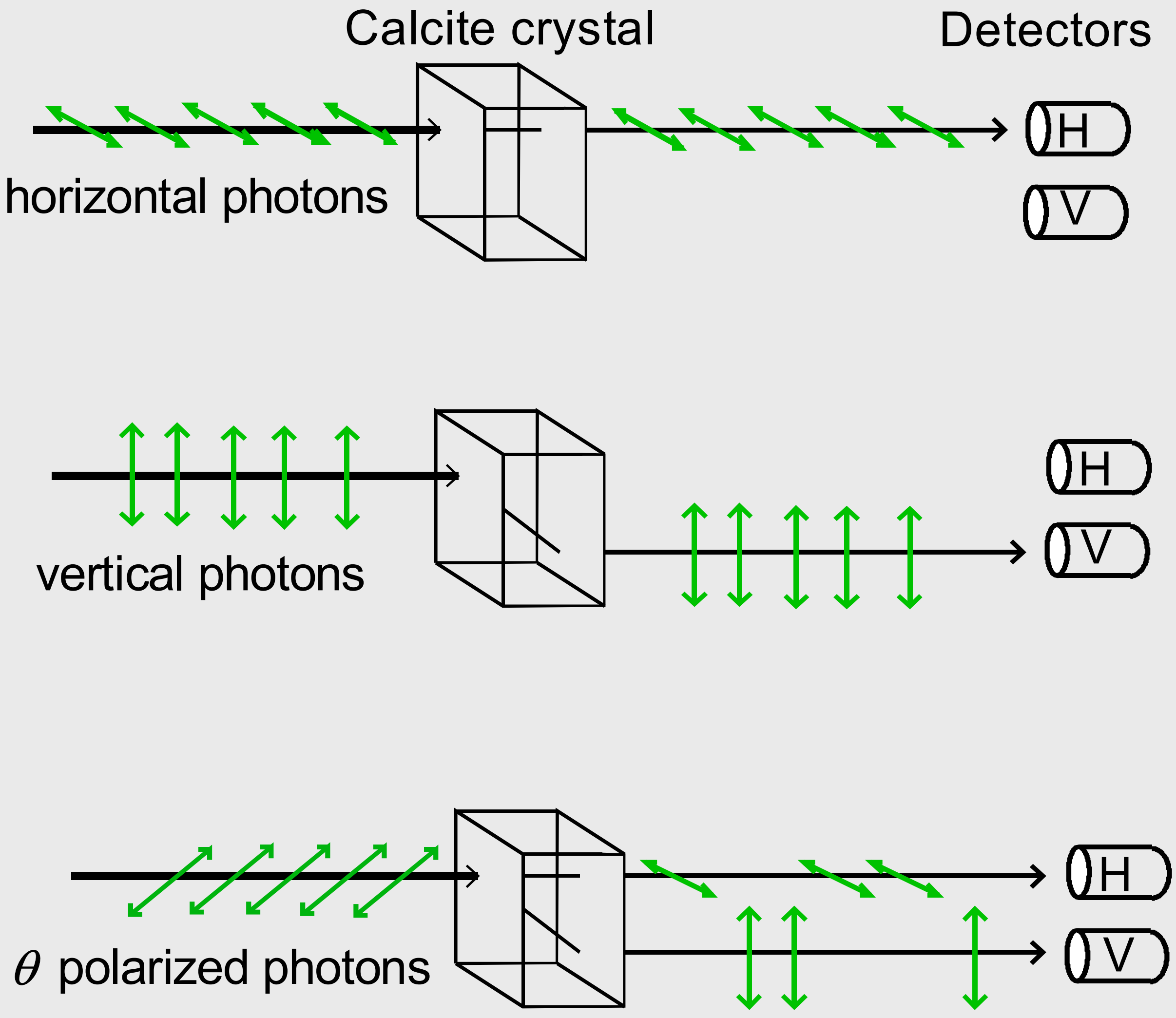
the Superposition Principle:

- Between any two reliably distinguishable states of a physical system (for example vertically and horizontally polarized single photons) there are intermediate states (for example diagonal photons) that are not reliably distinguishable from either original state

- The possible physical states correspond to directions in space—not ordinary 3-dimensional space, but an *n*-dimensional space where *n* is the system's maximum number of reliably distinguishable states.

- Any direction is a possible state, but two states are reliably distinguishable if only if their directions are perpendicular.

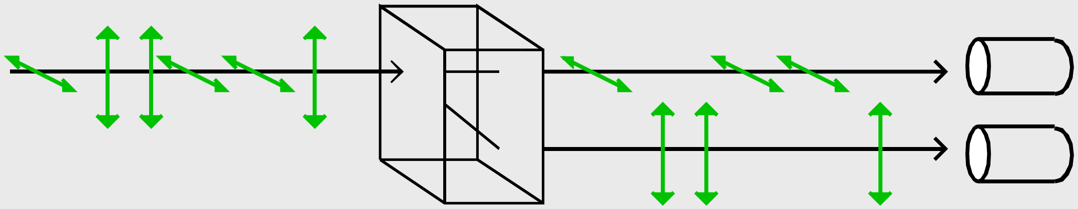
Using Polarized Photons to Carry Information


Calcite crystal
Detectors
horizontal photons
H
V
vertical photons
H
V
θ polarized photons
H
V
probability cos² θ
probability sin² θ

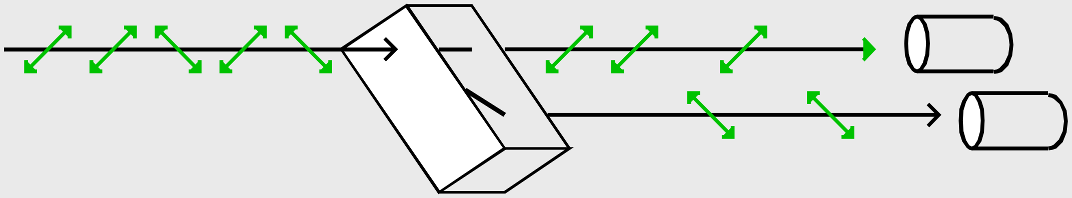
Photons behave reliably if measured along an axis parallel or perpendicular to their original polarization. Used in this way, each photon can carry one reliable bit of information.

But measuring the photons along any other axis causes them to behave randomly, forgetting their original polarization direction.

A rectilinear (ie vertical vs horizontal) measurement distinguishes vertical and horizontal photons reliably, but randomizes diagonal photons.

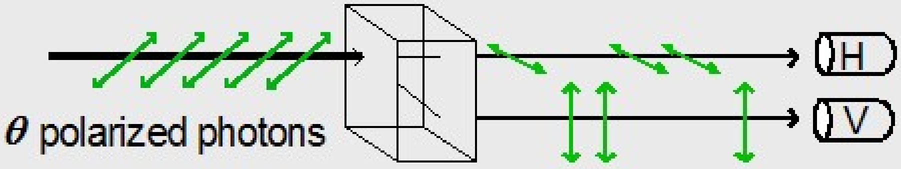


A diagonal measurement distinguishes diagonal photons reliably but randomizes rectilinear photons.



No measurement can distinguish all four kinds. This is not a limitation of particular measuring apparatuses, but a fundamental consequence of the uncertainty principle. This fundamental limitation gives rise to the possibility of quantum money and quantum cryptography.

Prof. William Wootters' pedagogic analogy for quantum measurement



Like a pupil confronting a strict teacher, a quantum system being measured is forced to choose among a set of distinguishable states (here 2) characteristic of the measuring apparatus.

Teacher: Is your polarization vertical or horizontal?

Pupil: Uh, I am polarized at about a 55 degree angle from horizontal.

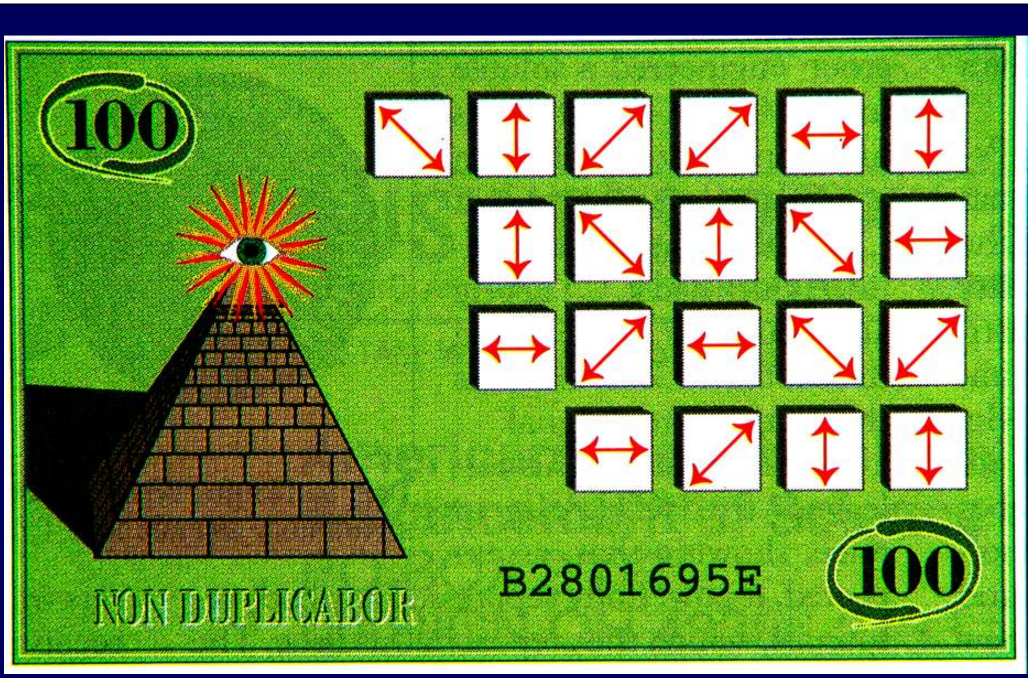
Teacher: I believe I asked you a question. Are you vertical or horizontal?

Pupil: Horizontal, sir.

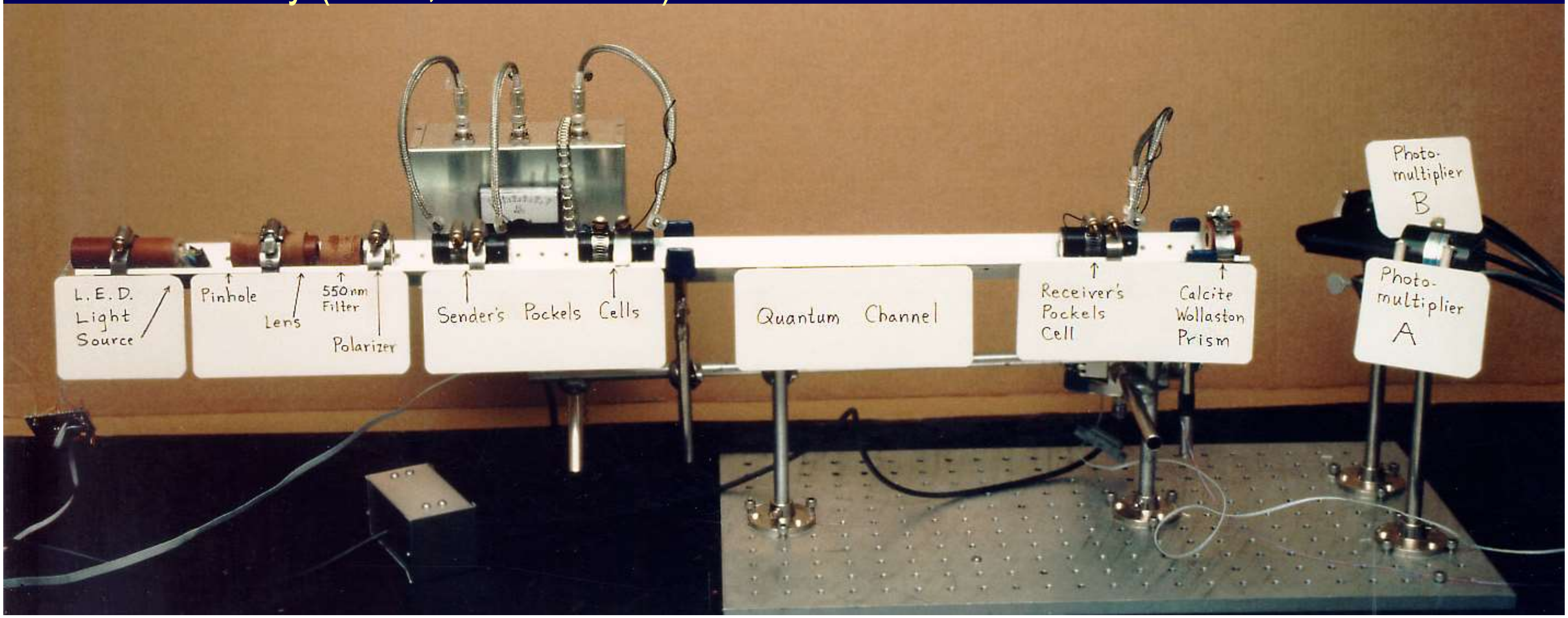
Teacher: Have you ever had any other polarization?

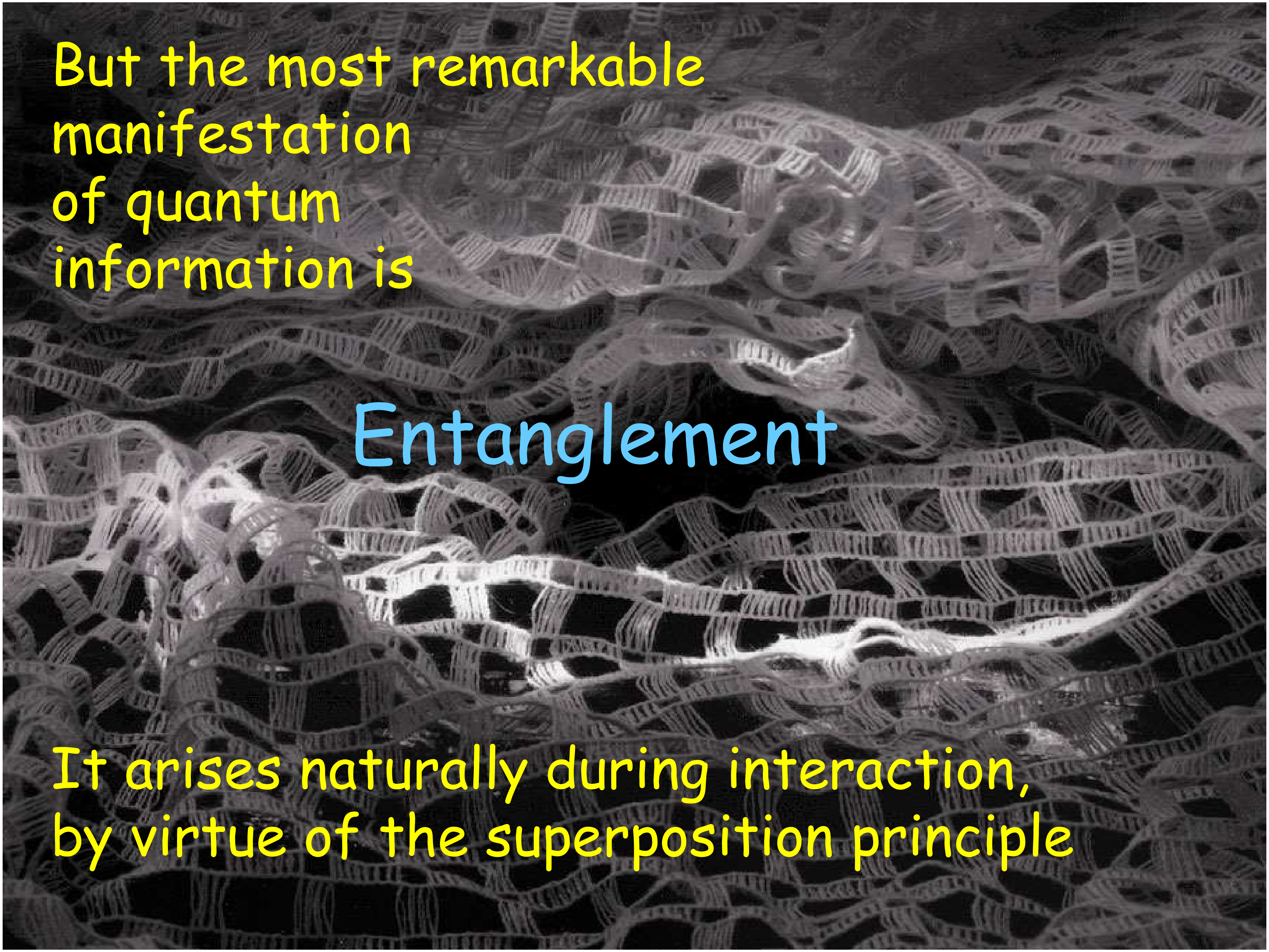
Pupil: No, sir. I was always horizontal.

Quantum money (Wiesner ’70, ’83) cannot be copied by a counterfeiter, but can be checked by the bank, which knows the secret sequence of polarized photons it should contain.



Quantum cryptography uses polar- ized photons to generate shared secret information between parties who share no secret initially (BB84, BBBSS92…)



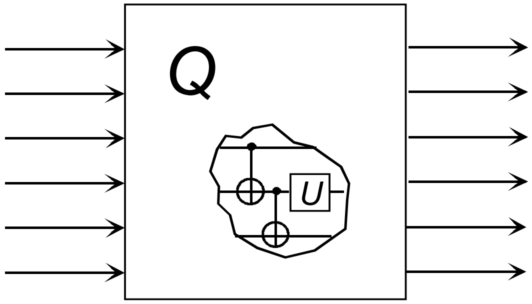


But the most remarkable manifestation of quantum information is

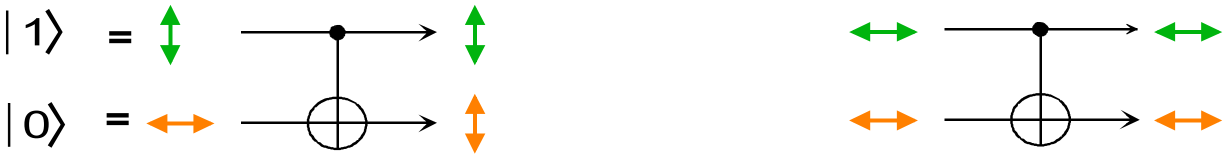
Entanglement

It arises naturally during interaction, by virtue of the superposition principle

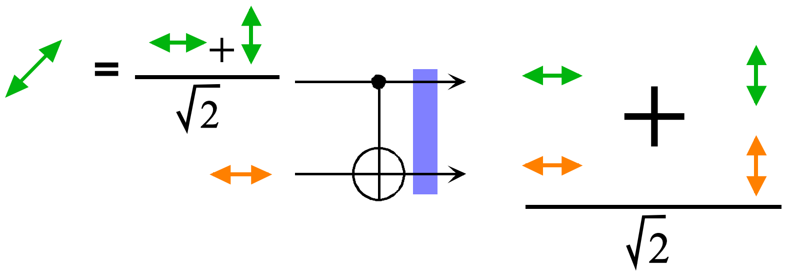
Any quantum data processing can be done by 1- and 2-qubit gates acting on qubits.



The 2-qubit XOR or "controlled-NOT" gate flips its 2nd input if its first input is 1, otherwise does nothing.

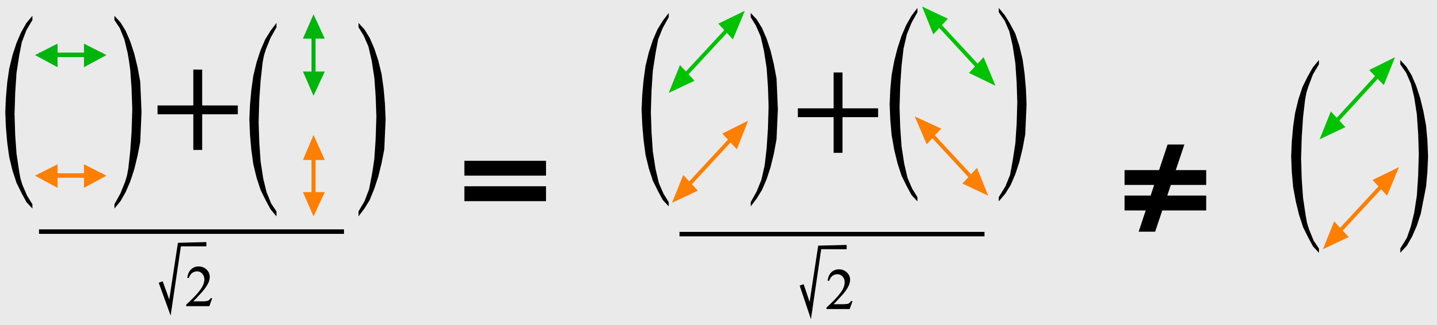


A superposition of inputs gives a superposition of outputs.



An entangled state

This entangled state of two photons behaves in ways that cannot be explained by supposing that each photon has a state of its own.



The two photons may be said to be in a definite state of *sameness* of polarization even though neither photon has a polarization of its own.

Entanglement sounds like a fuzzy new-age idea.

(In San Francisco in 1967, the “Summer of Love”, one often met people who felt they were in perfect harmony with one another, even though they had no firm opinions about anything.)

Wikiwatcher1 CC Share alike

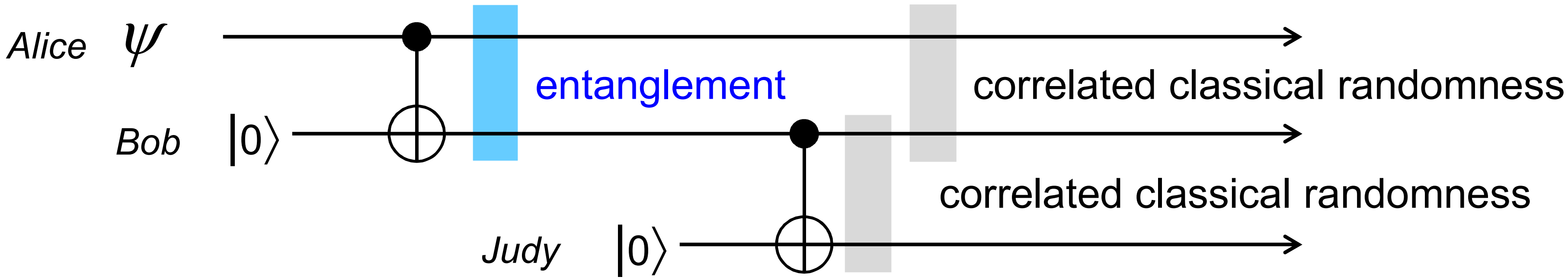
Hippies thought that with enough LSD, everyone could be in perfect harmony with everyone else.

Now we have a quantitative theory of entanglement and know it is *monogamous*: the more entangled two systems are with each other, the less entangled they can be with anything else.

The Monogamy of Entanglement

• If A and B are maximally entangled with each other, they can't they be entangled with anyone else.
• If one member of an entangled pair tries to share the entanglement with a third party, each pairwise relation is reduced to mere correlated randomness.

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


Alice
ψ
Bob
$|0\rangle$
entanglement
correlated classical randomness
correlated classical randomness
Judy
$|0\rangle$

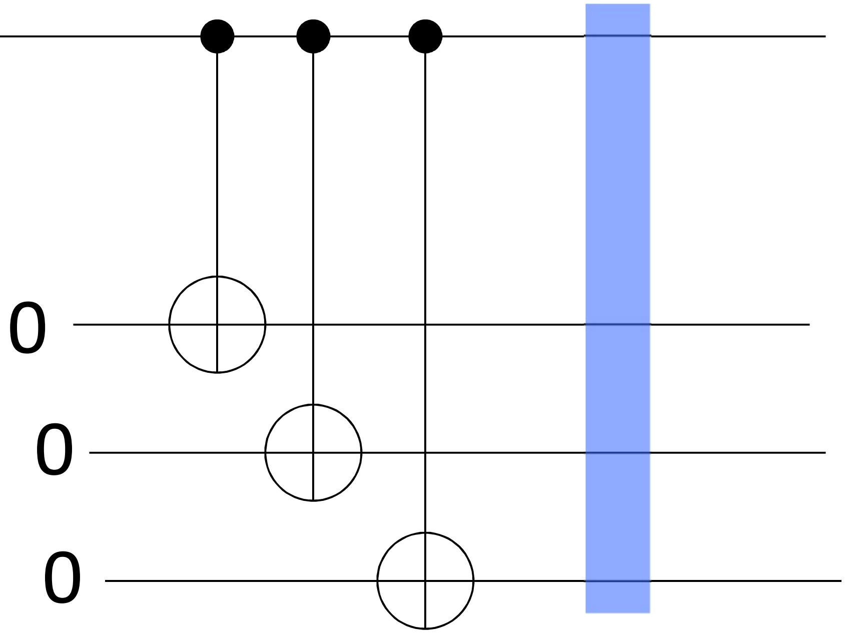
If one of Bob's girlfriends leaves the scene, Bob will find his relationship with the other reduced to mere correlated randomness. If they both stick around, he ends up perfectly entangled, not with either one, but with the now nontrivial *relationship* between them, an appropriate punishment.

How entanglement hides, creating a classical-appearing world

ψ

System

Parts of the system's low-entropy environment e.g. photons from the sun



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.

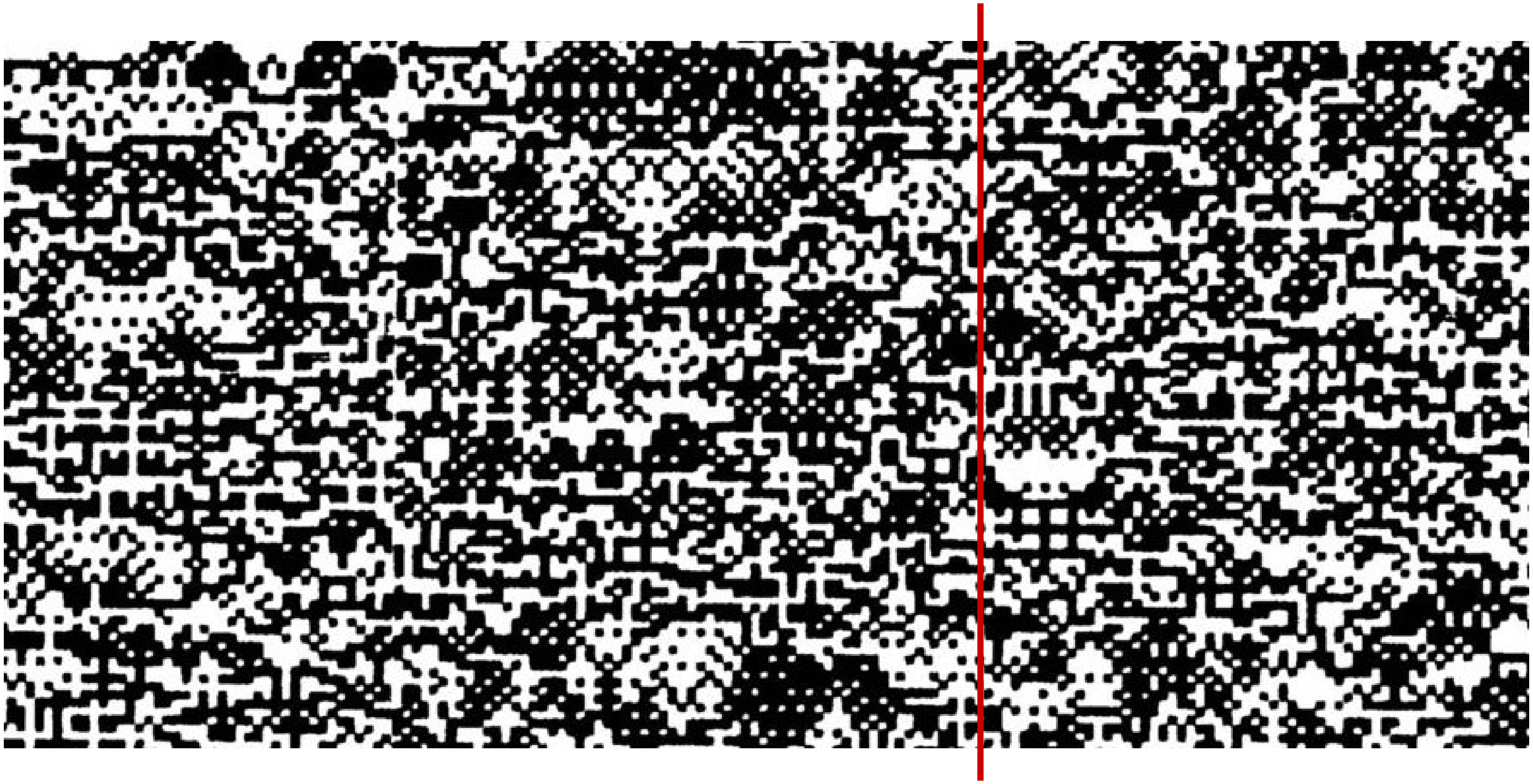
(This typically happens when the environment is not at thermal equilibrium, and contains many subsystems that interact in a commuting fashion with the system but not with each other. The earth's environment is like that.)

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.

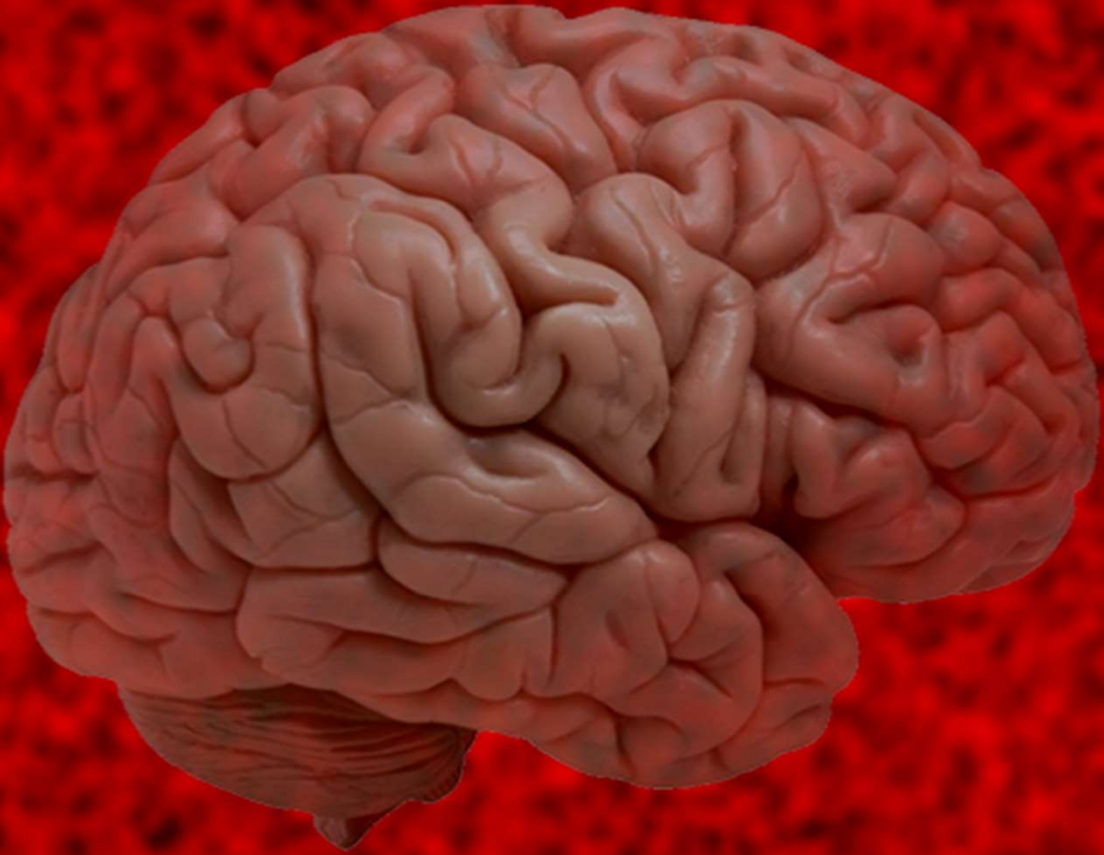
But if the reflected photons were instead trapped inside a reflective box, they would be repeatedly absorbed and re-emitted from the Earth, obfuscating the former redundant correlations as the system equilibrates, and rendering the system no longer classical.

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.

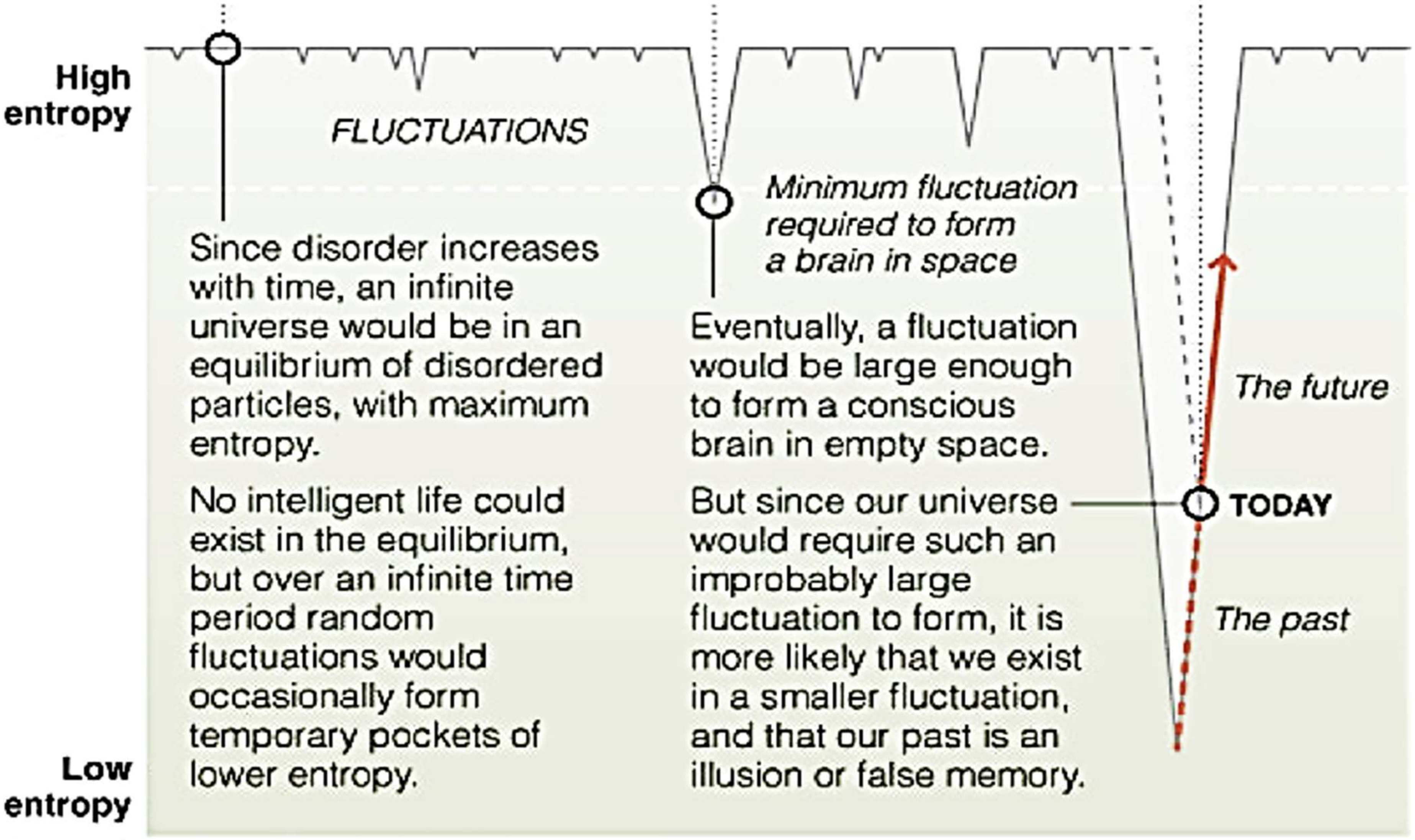


Conversely, a deep structure, one that seems to have had a long history, might just be the result of an unlikely thermal fluctuation, a so-called Boltzmann Brain.

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.***



Minimum fluctuation required to form a brain in space

Since disorder increases with time, an infinite universe would be in an equilibrium of disordered particles, with maximum entropy.

No intelligent life could exist in the equilibrium, but over an infinite time period random fluctuations would occasionally form temporary pockets of lower entropy.

Eventually, a fluctuation would be large enough to form a conscious brain in empty space.

But since our universe would require such an improbably large fluctuation to form, it is more likely that we exist in a smaller fluctuation, and that our past is an illusion or false memory.

Source: Sean Carroll, California Institute of Technology

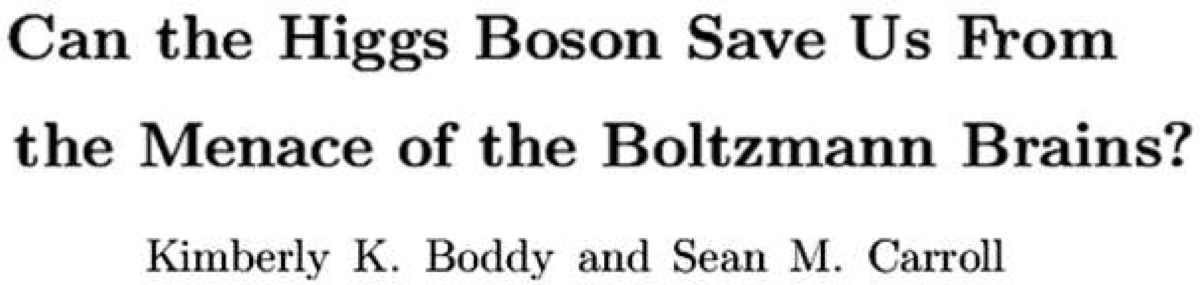
JONATHAN CORUM/THE NEW YORK TIMES

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

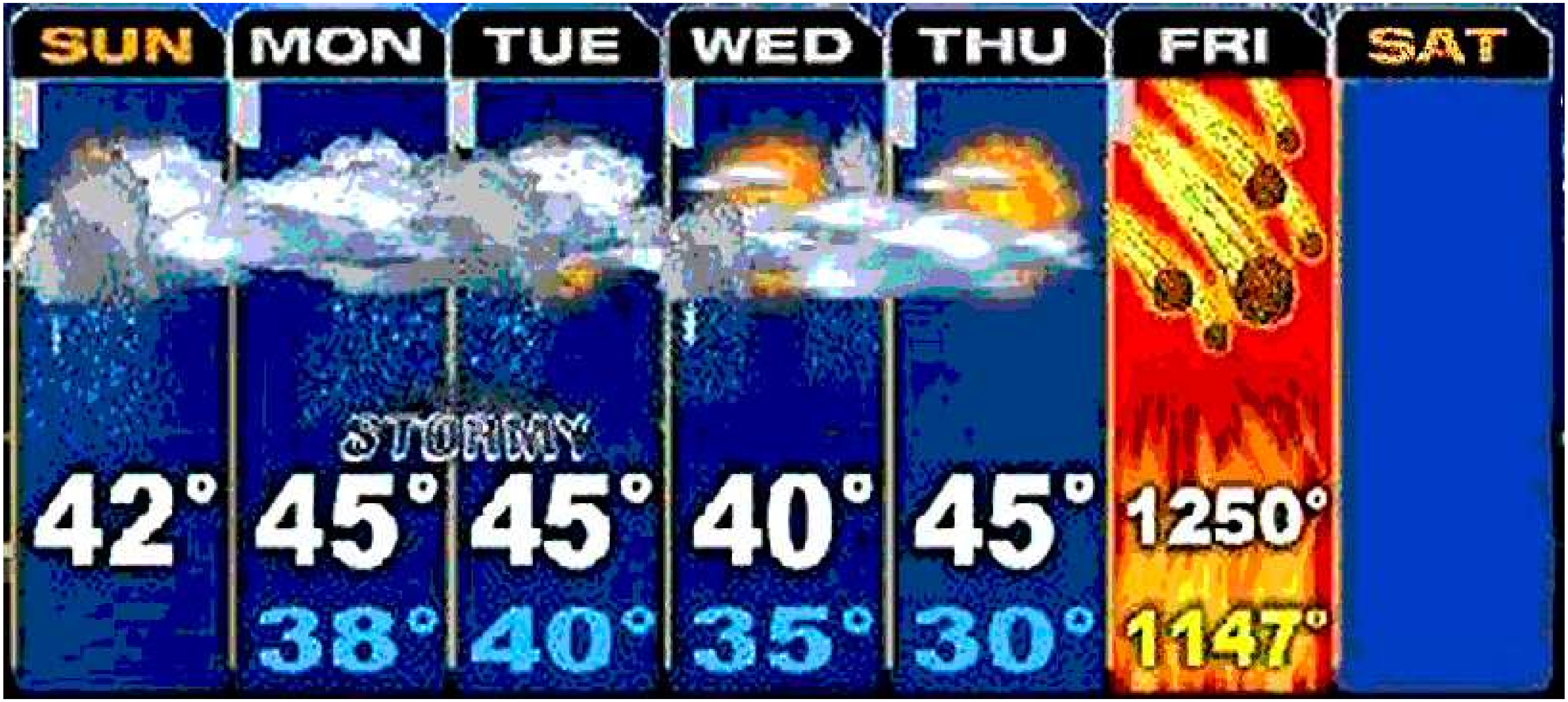
e.g. arxiv:1308.4686



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

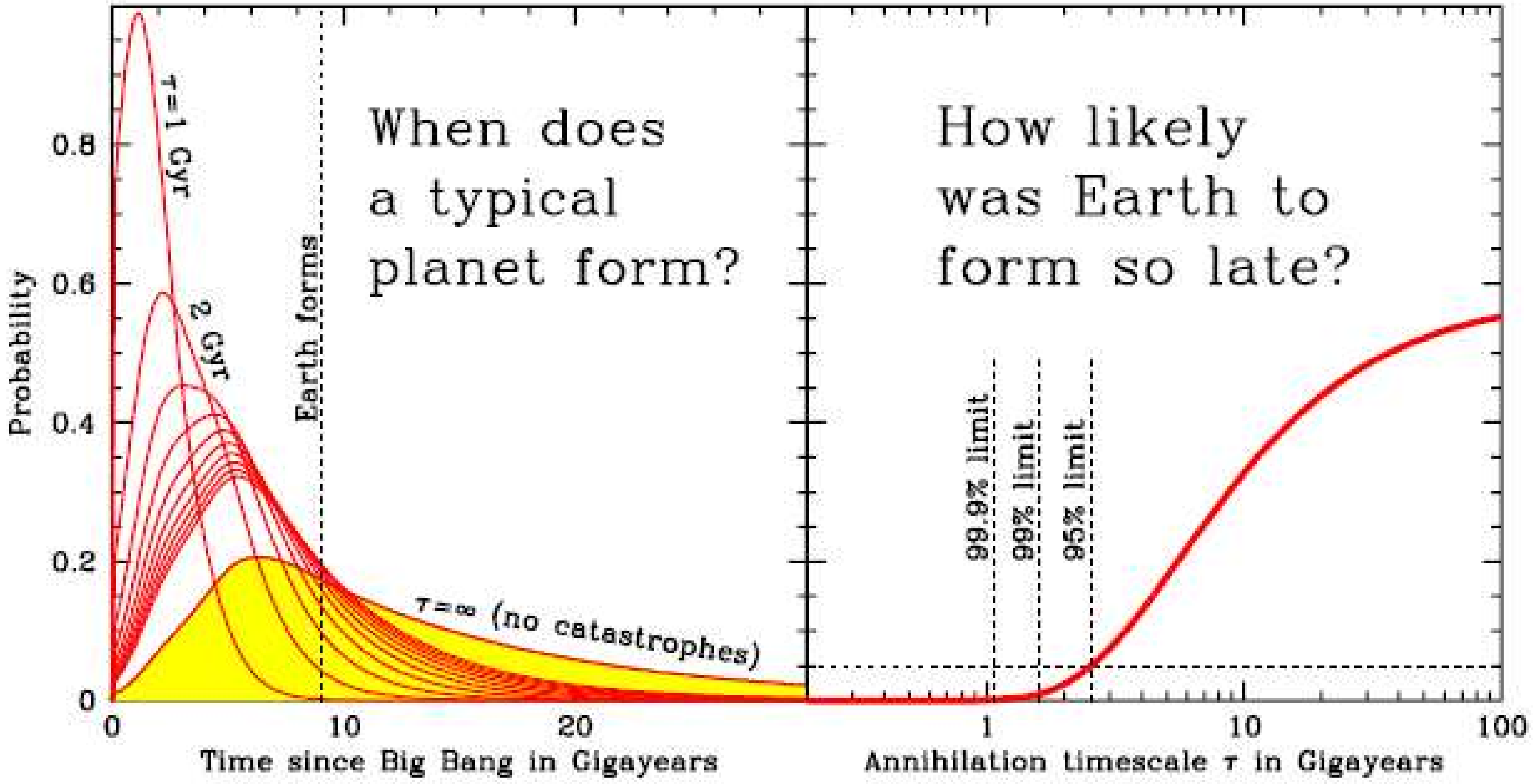
Kimberly K. Boddy and Sean M. Carroll

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.



Five 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.



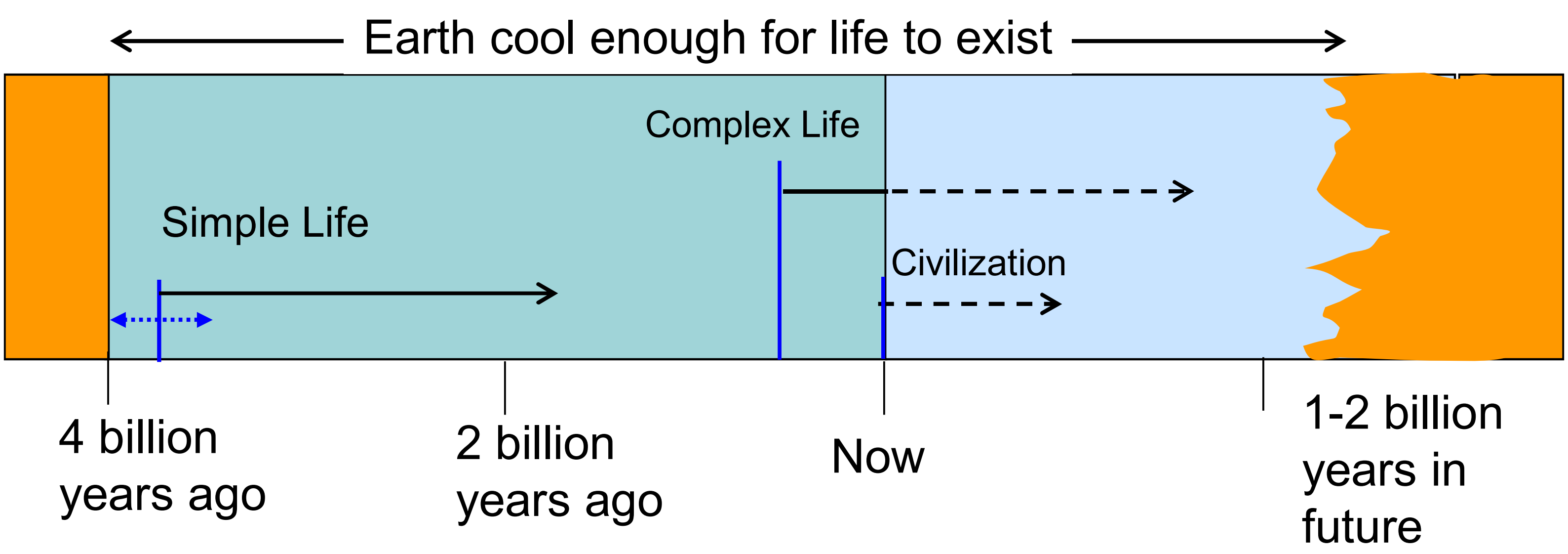
Doomsday arguments illustrate undisciplined thinking based on assumed typicality of the observer, without considering ways in which the observer may be atypical.

“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.

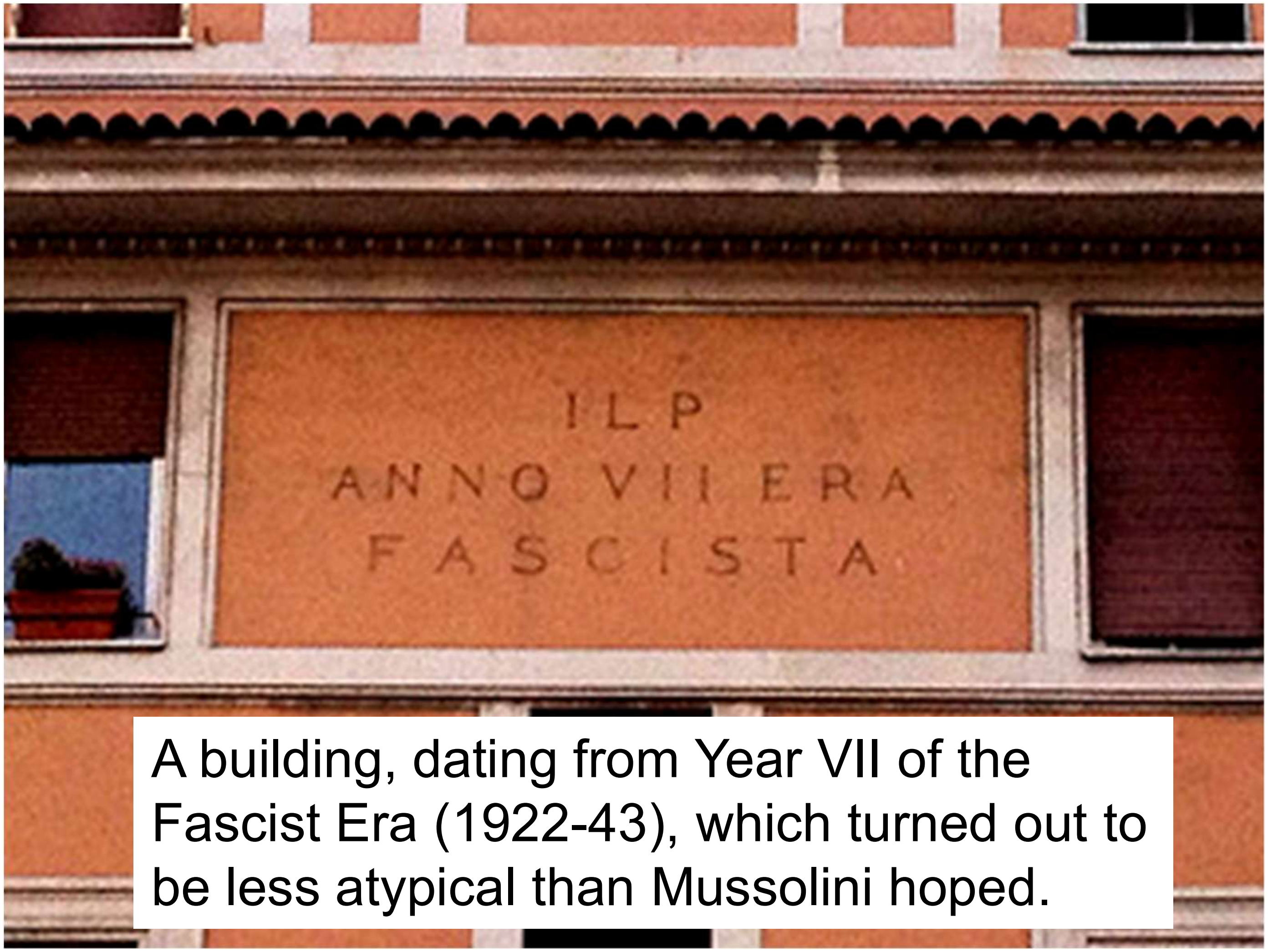
A more severe doomsday question occurs in connection with *civilization*, which has existed only a few *millionths* of the time potentially available for it (e.g. before the sun gets too hot).



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?

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, 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 remote workshop talk)

Wigner’s Friend

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

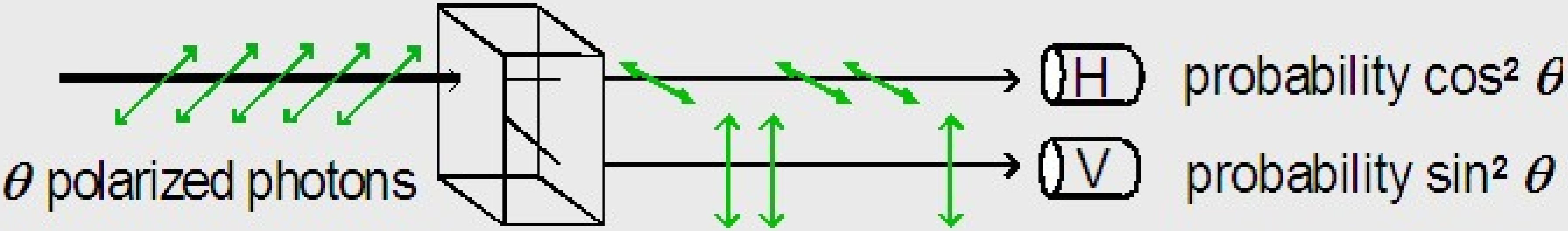
Eugene Wigner imagined a gentler experiment, relevant to the Quantum Boltzmann Brain problem:

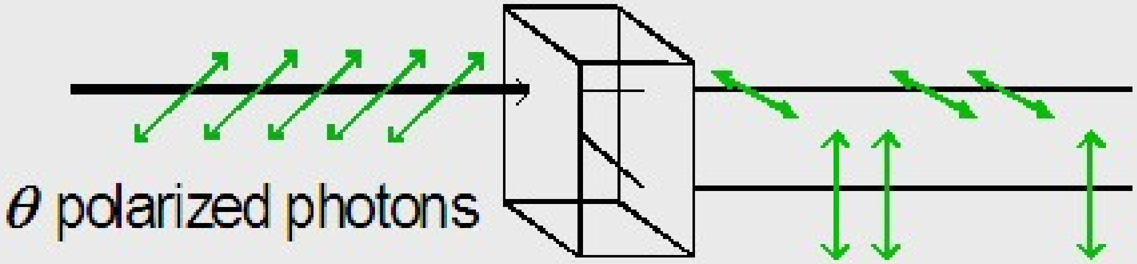
Wigner’s friend performs a quantum measurement with two outcomes but only tells Wigner what happened later.

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.

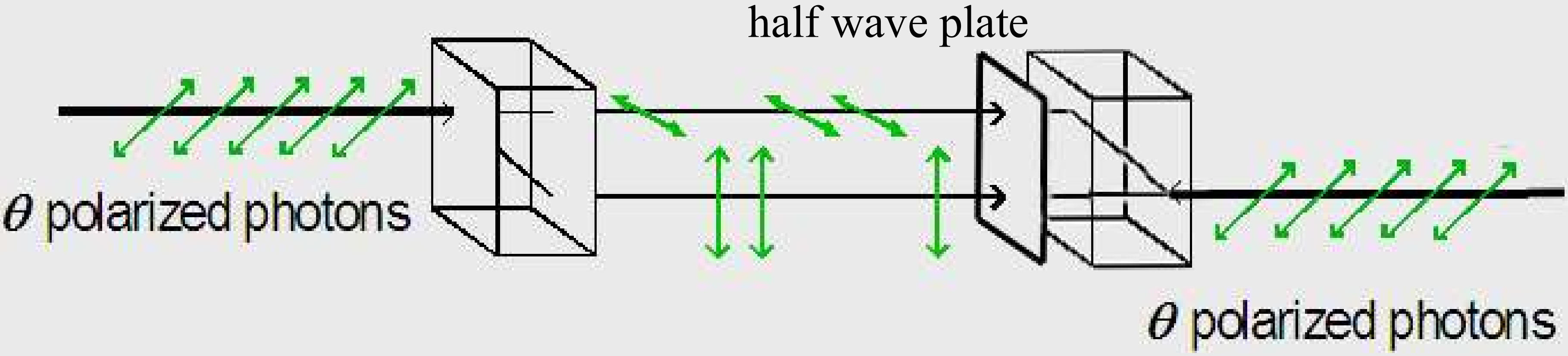
In principle (and even in practice, for atom-sized friends) Wigner can contrive for the friend to undo the measurement and forget its result—a “quantum eraser” experiment.

Entanglement and the origin of Quantum Randomness





If no one observes the photons, their random “behavior” can be undone.



Metaphorically speaking, it is the public embarrassment of the pupil, in front of the whole class, that makes him forget his original polarization.

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.

Open questions

- Wigner’s Friend’s experiences, if any

- Do entanglement and topological order enable generic fault-tolerant memory and self-organization at equilibrium (escape from Gibbs phase law)

- Are there cosmologies (e.g. eternal inflation) providing perpetual disequilibrium sufficient to support unbounded fault-tolerant classical self-organization

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 wavefunction, 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

Extra slides

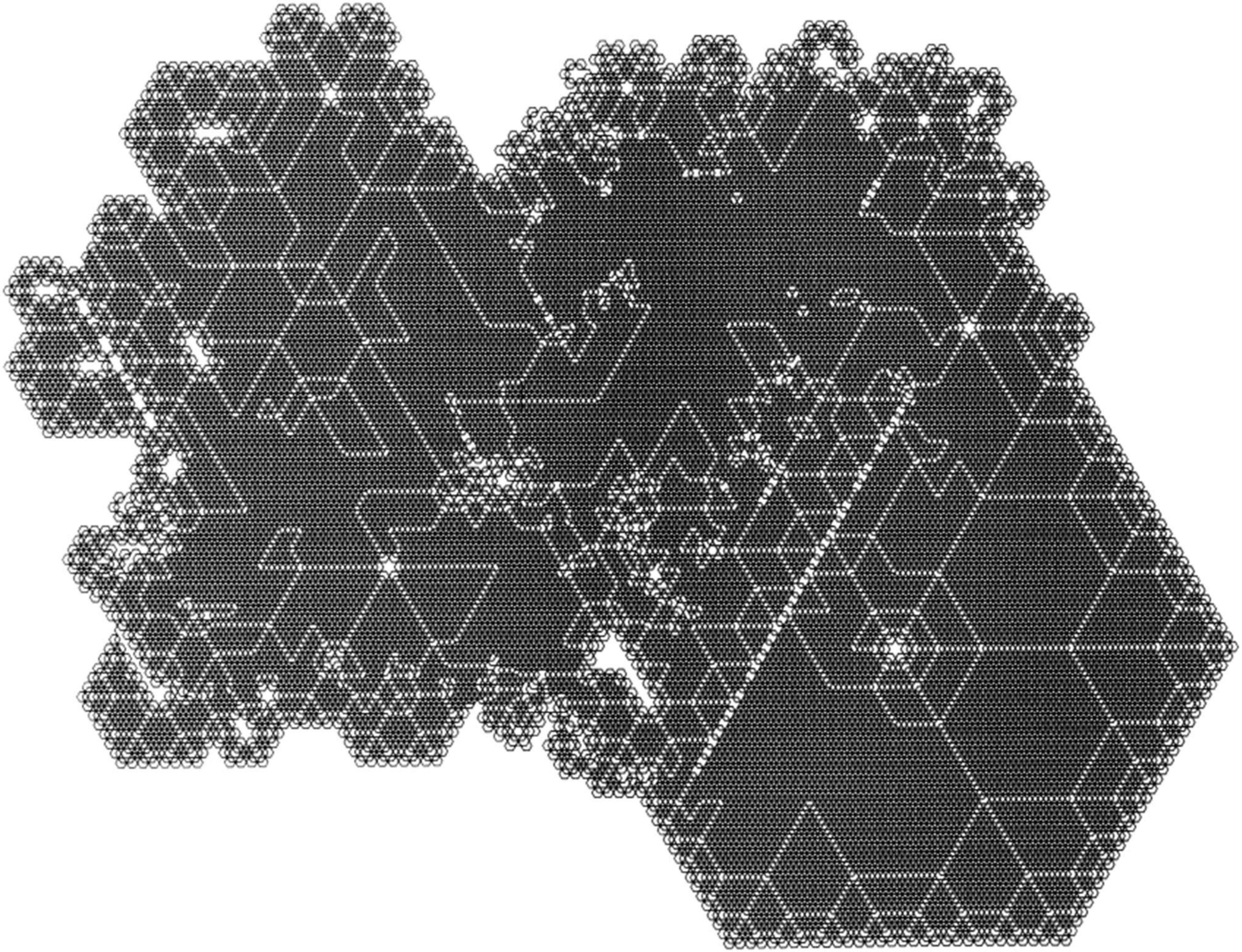
To make the quantitative definition of logical depth more stable with respect small variations of the string *x* and the universal machine *U,* the definition needs to be refined to take weighted account of all programs for computing the object, not just the smallest.

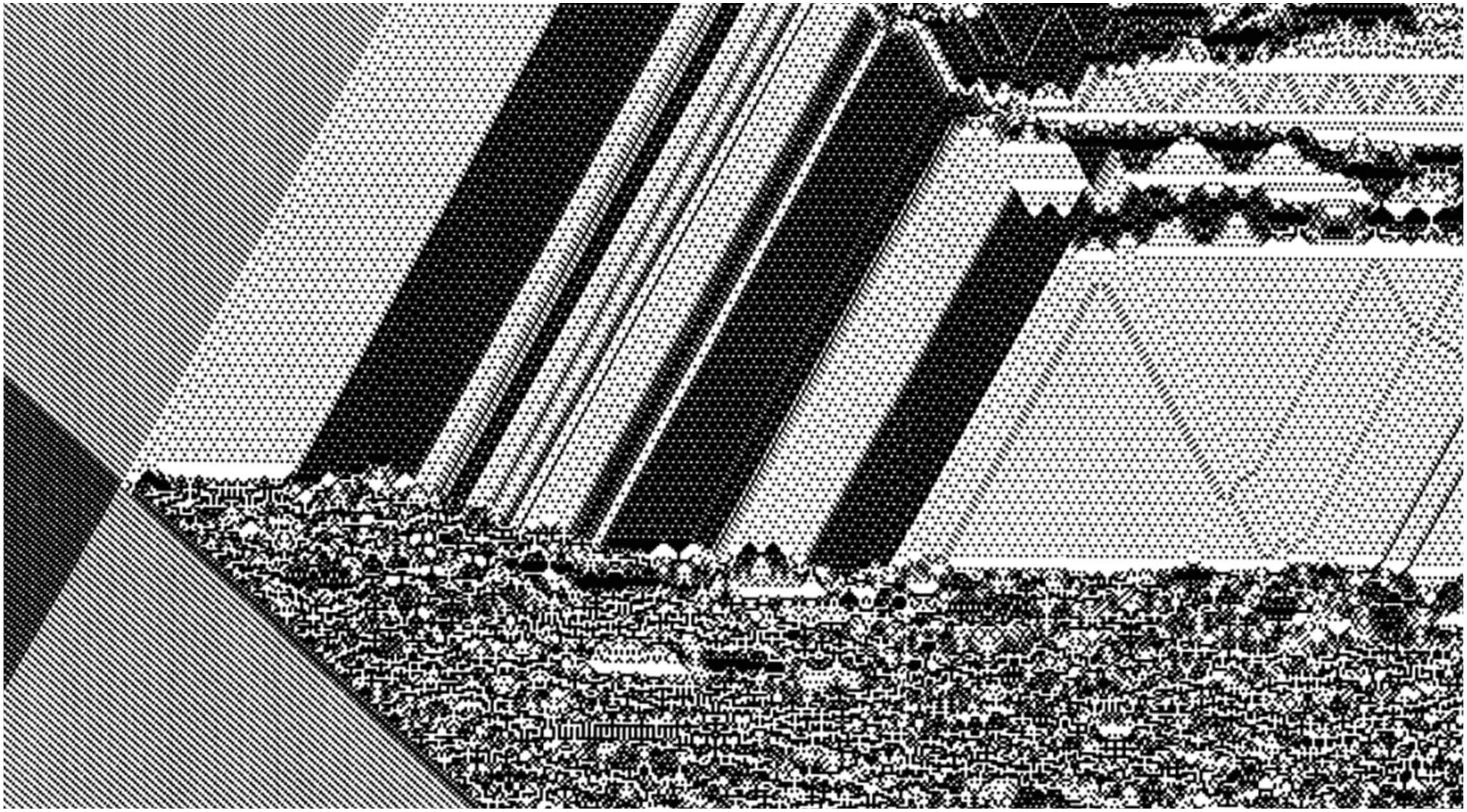
The *s*-significant depth of a string *x,* denoted $D_s(x)$, is defined as the least run time of any *s*-incompressible program to compute *x*:

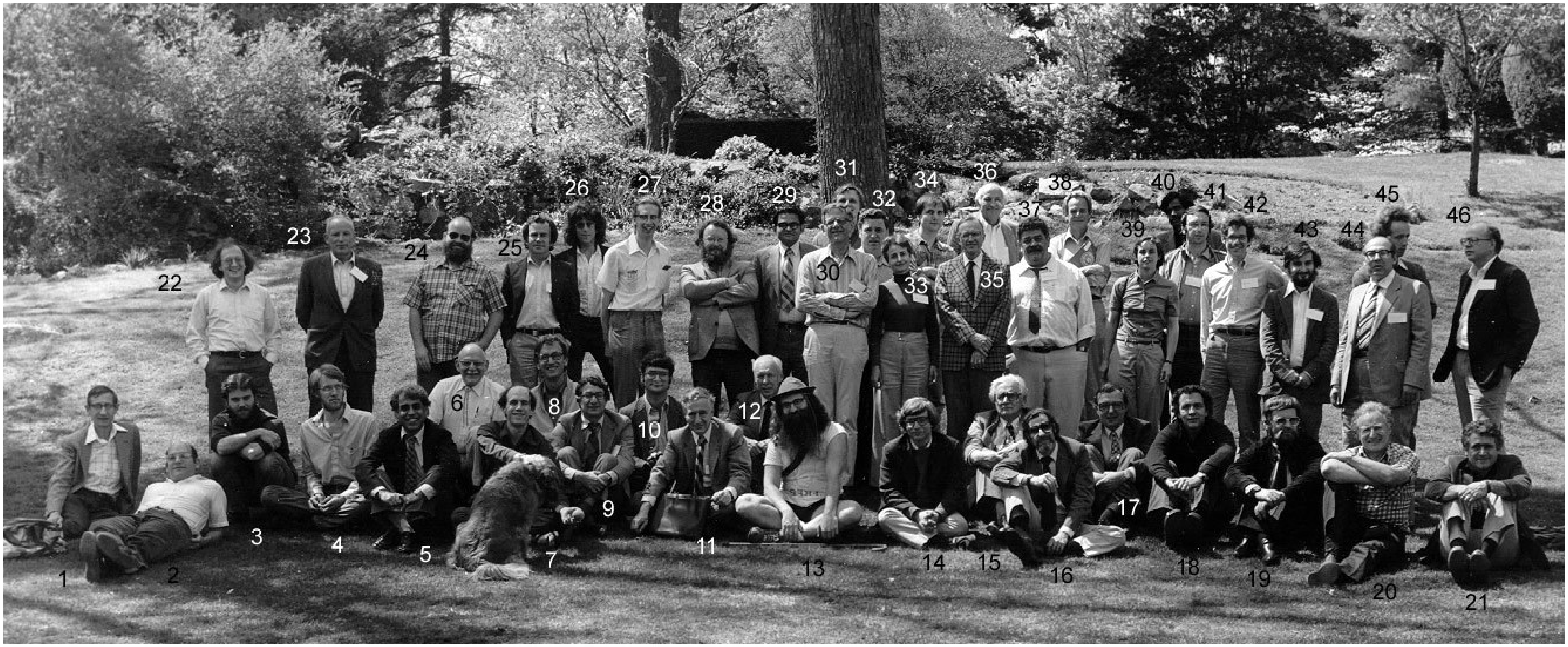
$$D_s(x) = \min\{T(p): U(p)=x \;\&|p|-|p^*|<s\}.$$

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.







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

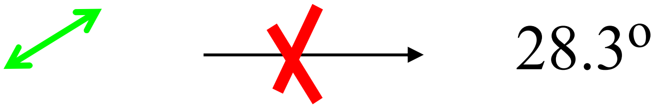
1 Freeman Dyson
2 Gregory Chaitin
3 James Crutchfield
4 Norman Packard
5 Panos Ligomenides
6 Jerome Rothstein
7 Carl Hewitt
8 Norman Hardy
9 Edward Fredkin
10 Tom Toffoli
11 Rolf Landauer
12 John Wheeler

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

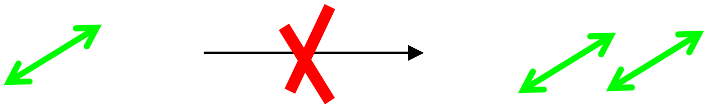
25 Robert Suaya
26 Stan Kugell
27 Bill Gosper
28 Lutz Priese
39 Madhu Gupta
30 Paul Benioff
31 Hans Moravec
32 Ian Richards
33 Marian Pour-El
34 Danny Hillis
35 Arthur Burks
36 John Cocke

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

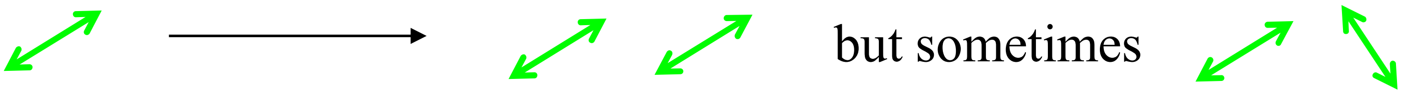
Measuring an unknown photon’s polarization exactly is impossible (no measurement can yield more than 1 bit about it).

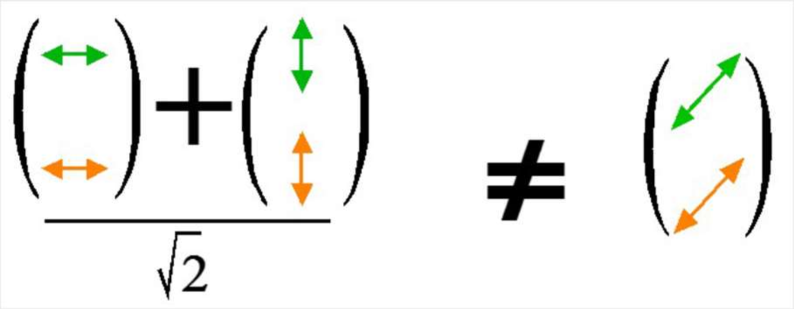


Cloning an unknown photon is impossible. (If either cloning or measuring were possible the other would be also).



If you try to amplify an unknown photon by sending it into an ideal laser, the output will be polluted by just enough noise (due to spontaneous emission) to be no more useful than the input in figuring out what the original photon’s polarization was.





To see that the entangled state on the left is different from the two-diagonal-photon state on the right, do some simple algebra, using a more compact notation in place of the double-headed arrows.

First consider single photon states, which live in a 2-dimensional space. Let H represent a horizontal photon and V a vertical photon.

Then (H+V)/√2 is a 45 degree diagonal photon and (H-V)/√2 is a 135 degree diagonal photon. For states of two photons, one green and one orange, it is clear that the four states HH, HV, VH, and VV are all distinguishable (e.g. HH and HV can be distinguished by measuring the orange photon), so the 2-photon states live in a four-dimensional space. The entangled 2-photon state on the left is (HH+VV)/√2, a certain direction in four-dimensional space, while the 2-diagonal-photon state on the right,

((H+V)/√2) ((H+V)/√2) = (HH+HV+VH+VV)/2

is a *different* direction in four-dimensional space. Only the latter can be described by attributing a definite polarization to each photon.

Pedagogic analog of entanglement:

Twin pupils Remus and Romulus, who are each completely ignorant of all subjects, answering every question randomly, but they always give the same answer, even when questioned separately.

Teacher A: Remus, what color is growing grass?

Remus: Pink, sir.

Teacher B (in another classroom): Romulus, what color is growing grass?

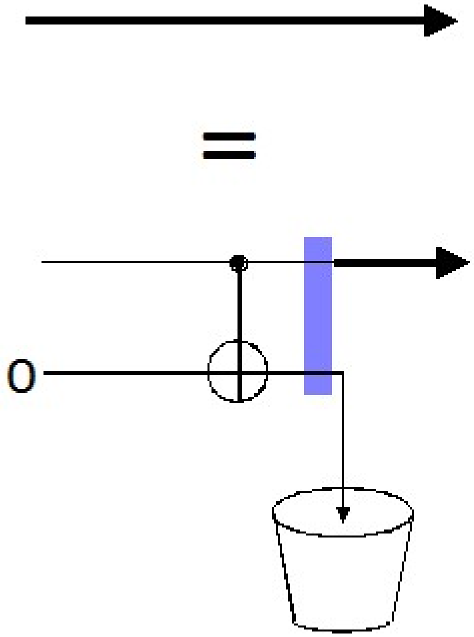
Romulus: Pink, ma’am.

Expressing Classical Data Processing in Quantum Terms

A Classical Bit is a qubit with one of the Boolean values 0 or 1

A classical wire is a quantum channel that conducts 0 and 1 faithfully but randomizes superpositions of 0 and 1.

This happens because the data passing through the wire interacts with its environment, causing the environment to acquire a copy of it, if it was 0 or 1, and otherwise become entangled with it.



wastebasket symbolizes loss of
Information into the environment

A classical channel is a quantum channel with an eavesdropper.

A classical computer is a quantum computer handicapped by having eavesdroppers on all its wires.

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 ρ of this patch can be expressed as a weighted sum

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

  where σ is a thermal state “depleted” in π_{BB} .
- An all-powerful Preparator tosses a λ-biased coin, and prepares π_{BB} or σ 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 ρ to have nonzero participation of a Boltzmann brain state π_{BB}, namely

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

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

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

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.