ÖAW AUSTRIAN ACADEMY OF SCIENCES

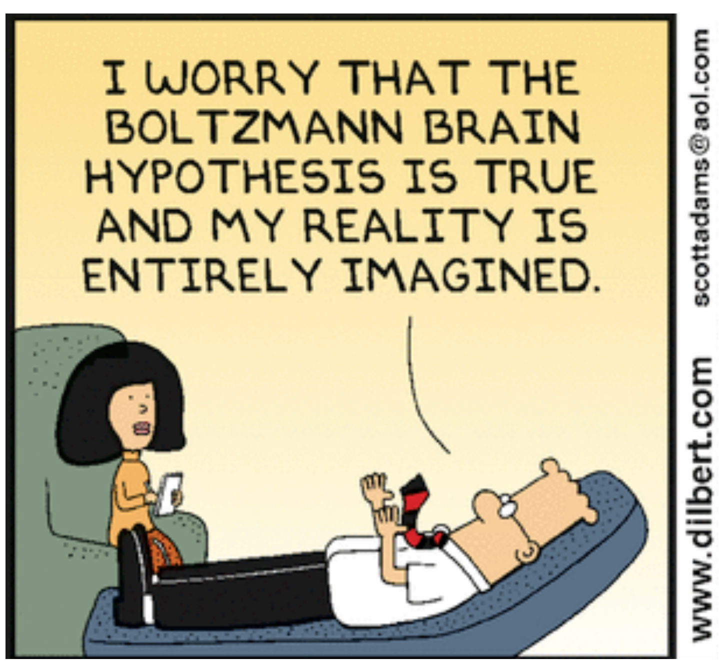
IQOQI – INSTITUTE FOR QUANTUM OPTICS AND QUANTUM INFORMATION VIENNA

Information-theoretic idealism

Markus P. Müller

Institute for Quantum Optics and Quantum Information, Vienna
Perimeter Institute for Theoretical Physics, Waterloo, Canada

Figur 1.



FQXi FOUNDATIONAL QUESTIONS INSTITUTE

Summary

I argue that our "standard" view of
the physical world may be ~~**wrong**~~.

only approximately true.

To solve some important foundational *and practical* problems,
it helps to take a (sort of) **idealist** approach.

Summary

I argue that our “standard” view of
the physical world may be ~~wrong~~.

only approximately true.

To solve some important foundational *and practical* problems,
it helps to take a (sort of) **idealist** approach.

idealism | ʌɪˈdɪəlɪz(ə)m, ʌɪˈdiːəlɪz(ə)m |
A diverse group of views that regard “mind” as primary, not matter.

Here: “mind”=**”pattern”:** mathematical, **information-theoretic** notion.
Irrelevant: consciousness, qualia, what we believe, want or feel.

Goal and structure of this talk

Goal: Convey the main idea and its methodological adequacy.

Goal and structure of this talk

Goal: Convey the main idea and its methodological adequacy.

Doing so, I will disregard

- most mathematical details (they are in the paper),
- most philosophical notions and issues, such as:

  Relation to realism, physicalism, Humean supervenience, historical predecessors, Carnap’s “Aufbau”, overlap with quantum interpretations, interpretations of probability, is it really idealism? (probably not), …

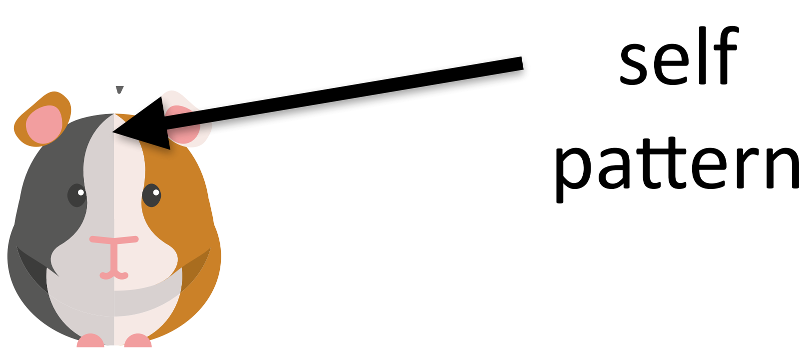
Goal and structure of this talk

Goal: Convey the main idea and its methodological adequacy.

Doing so, I will disregard

- most mathematical details (they are in the paper),
- most philosophical notions and issues, such as:
  Relation to realism, physicalism, Humean supervenience, historical predecessors, Carnap's "Aufbau", overlap with quantum interpretations, interpretations of probability, is it really idealism? (probably not), …

Colorful pictures, but notions **neither** inherently **human nor biological**.



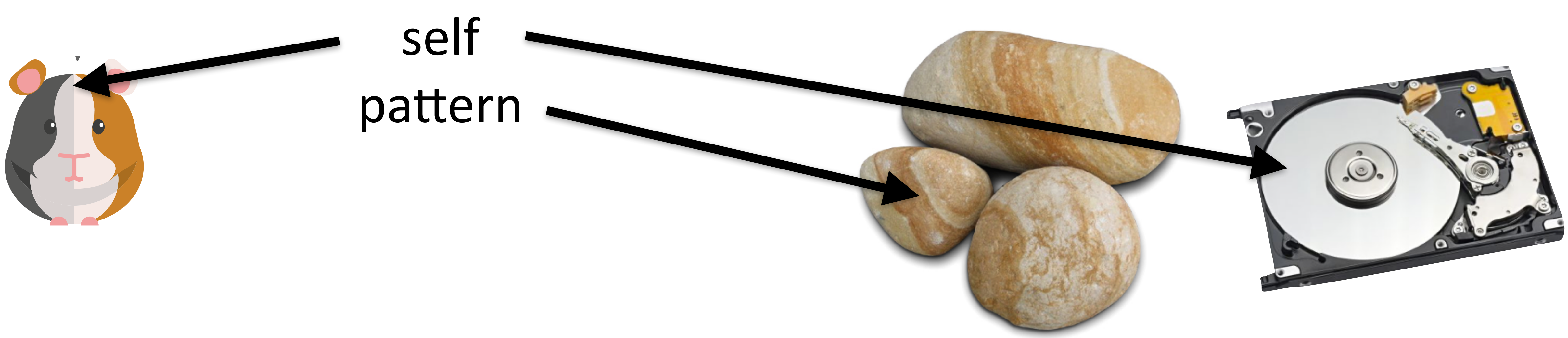
Goal and structure of this talk

Goal: Convey the main idea and its methodological adequacy.

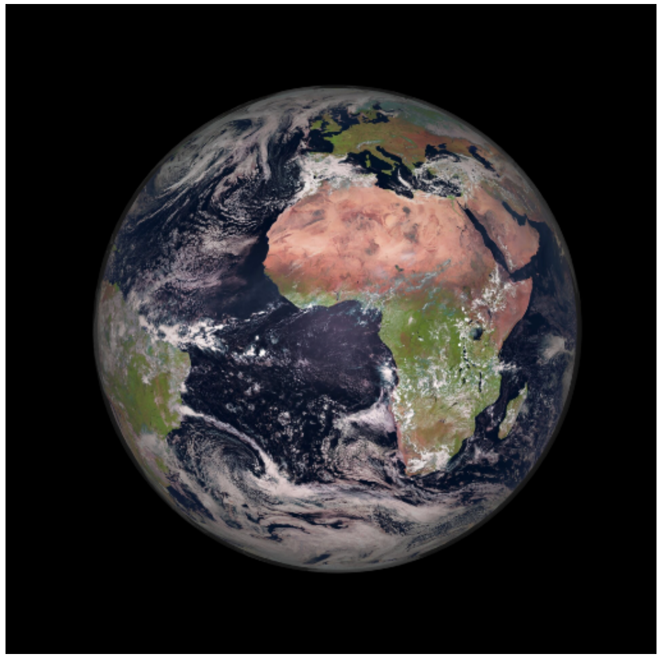
Doing so, I will disregard

- most mathematical details (they are in the paper),
- most philosophical notions and issues, such as:
  Relation to realism, physicalism, Humean supervenience, historical predecessors, Carnap's "Aufbau", overlap with quantum interpretations, interpretations of probability, is it really idealism? (probably not), …

Colorful pictures, but notions **neither** inherently **human nor biological**.

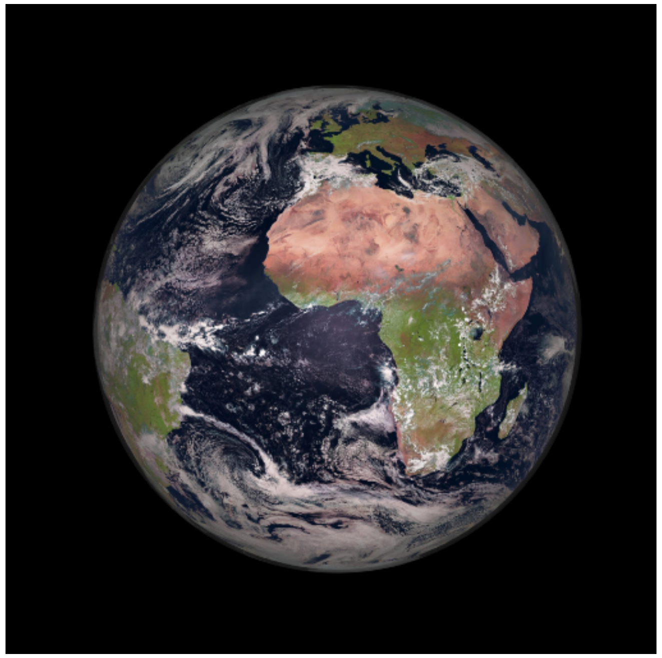


Motivating observation



Suppose we removed Earth from the universe, stopped its time evolution, and gave it to a team of infinitely smart and all-powerful scientists.

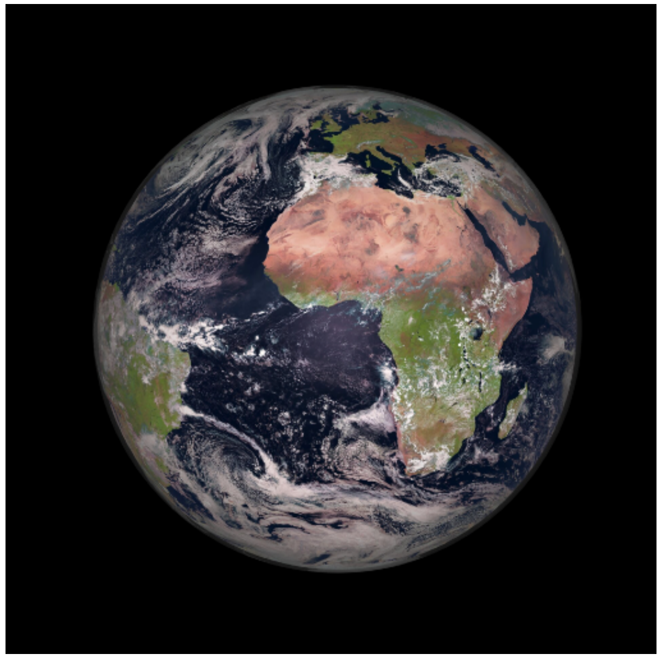
Motivating observation



Suppose we removed Earth from the universe, stopped its time evolution, and gave it to a team of infinitely smart and all-powerful scientists.

From the (classical) information contained in Earth, they could reconstruct an almost-perfect synthetic environment (universe) for it **with almost identical predictions for what happens to it in the future.**

Motivating observation



Suppose we removed Earth from the universe, stopped its time evolution, and gave it to a team of infinitely smart and all-powerful scientists.

From the (classical) information contained in Earth, they could reconstruct an almost-perfect synthetic environment (universe) for it **with almost identical predictions for what happens to it in the future.**

In a sense to be made precise (algorithmic information theory), *“a plausible rest of the universe”* can be seen as a **property** of (the information contained in) Earth.

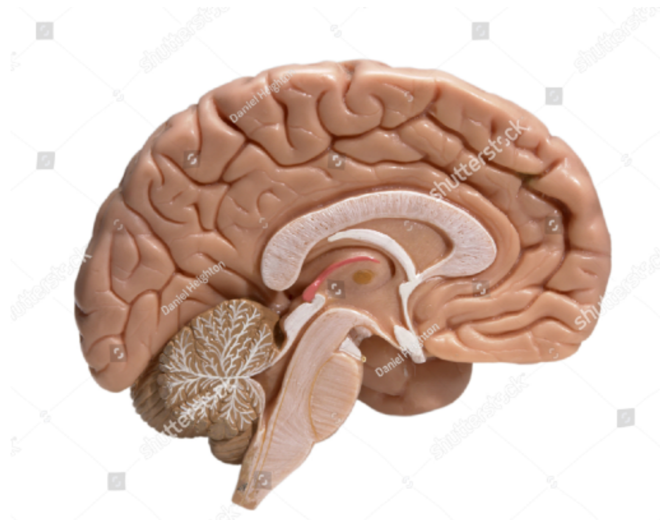
Motivating observation

Suppose we removed Earth from the universe, stopped its time evolution, and gave it to a team of infinitely smart and all-powerful scientists.

From the (classical) information contained in Earth, they could reconstruct an almost-perfect synthetic environment (universe) for it **with almost identical predictions for what happens to it in the future.**

In a sense to be made precise (algorithmic information theory), *"a plausible rest of the universe"* can be seen as a **property** of (the information contained in) Earth. **Same for smaller structures...**

Motivating observation



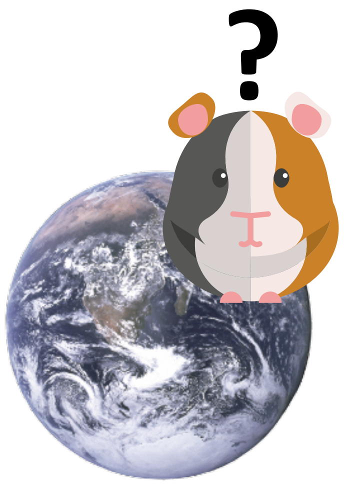
Suppose we removed Earth from the universe, stopped its time evolution, and gave it to a team of infinitely smart and all-powerful scientists.

From the (classical) information contained in Earth, they could reconstruct an almost-perfect synthetic environment (universe) for it **with almost identical predictions for what happens to it in the future.**

In a sense to be made precise (algorithmic information theory), *“a plausible rest of the universe”* can be seen as a **property** of (the information contained in) Earth. **Same for smaller structures…**

Outline

1. **Conceptual puzzles**
   … that challenge the standard view.



2. **Sketch of an idealist (toy) theory**
   … "self" fundamental, external world emergent.

3. **Objective reality as a emergent approximation**
   … probabilistic zombies, and other surprises.

4. **Example: dissolution of the Boltzmann brain problem**

Outline

1. **Conceptual puzzles**

   … that challenge the standard view.

2. **Sketch of an idealist (toy) theory**

   … "self" fundamental, external world emergent.

3. **Objective reality as a emergent approximation**

   … probabilistic zombies, and other surprises.

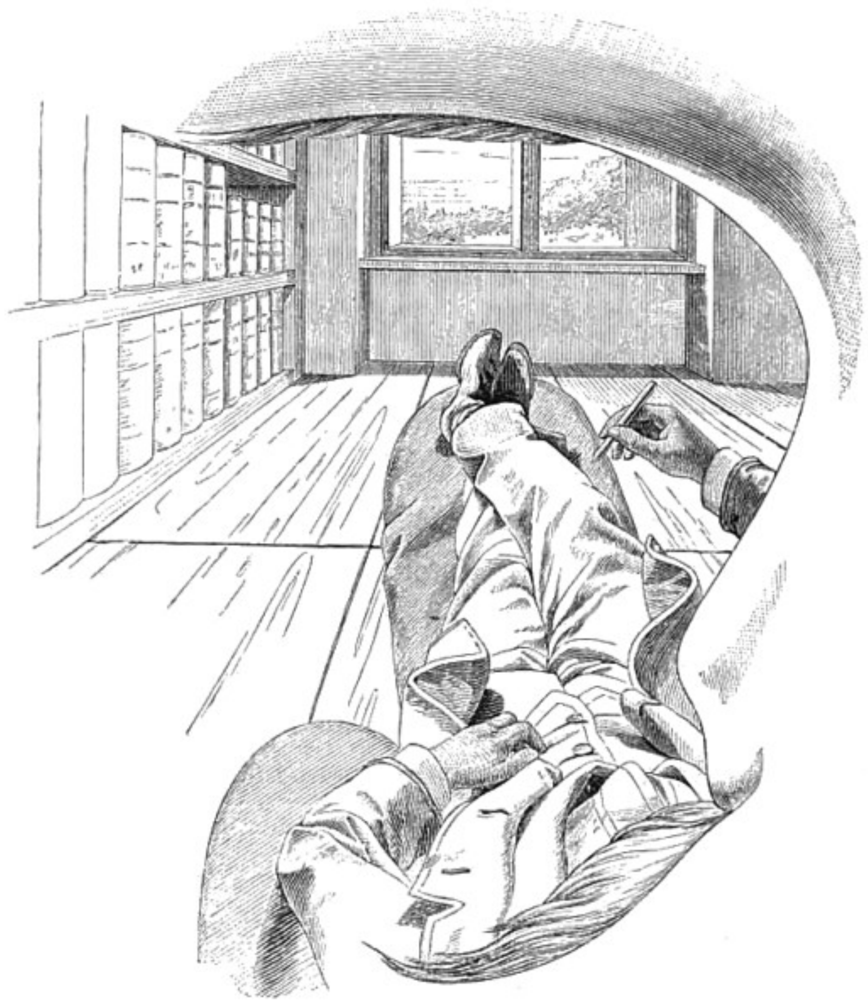
4. **Example: dissolution of the Boltzmann brain problem**

1. Conceptual puzzles

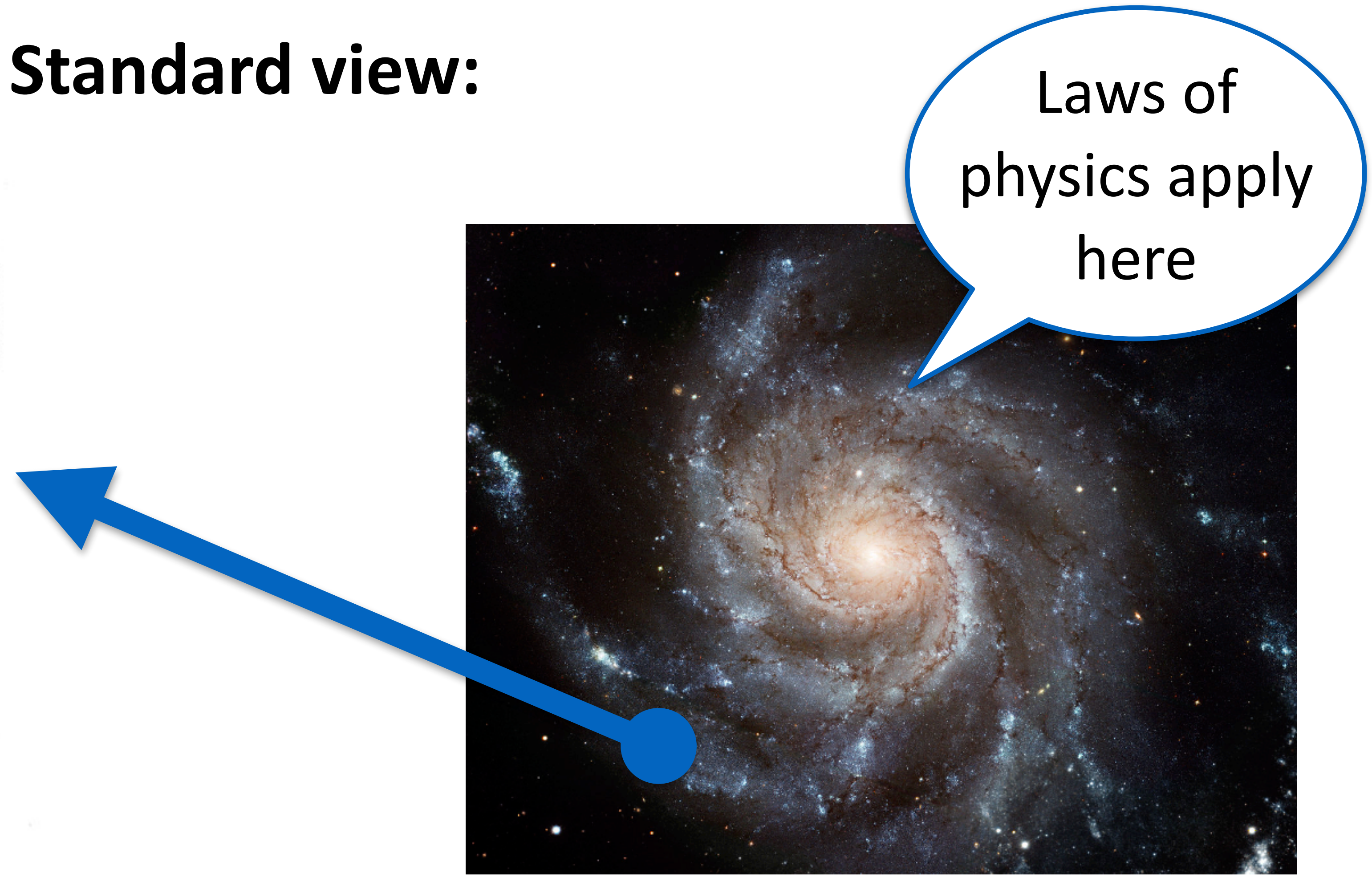
Standard view:

1. Conceptual puzzles

Standard view:

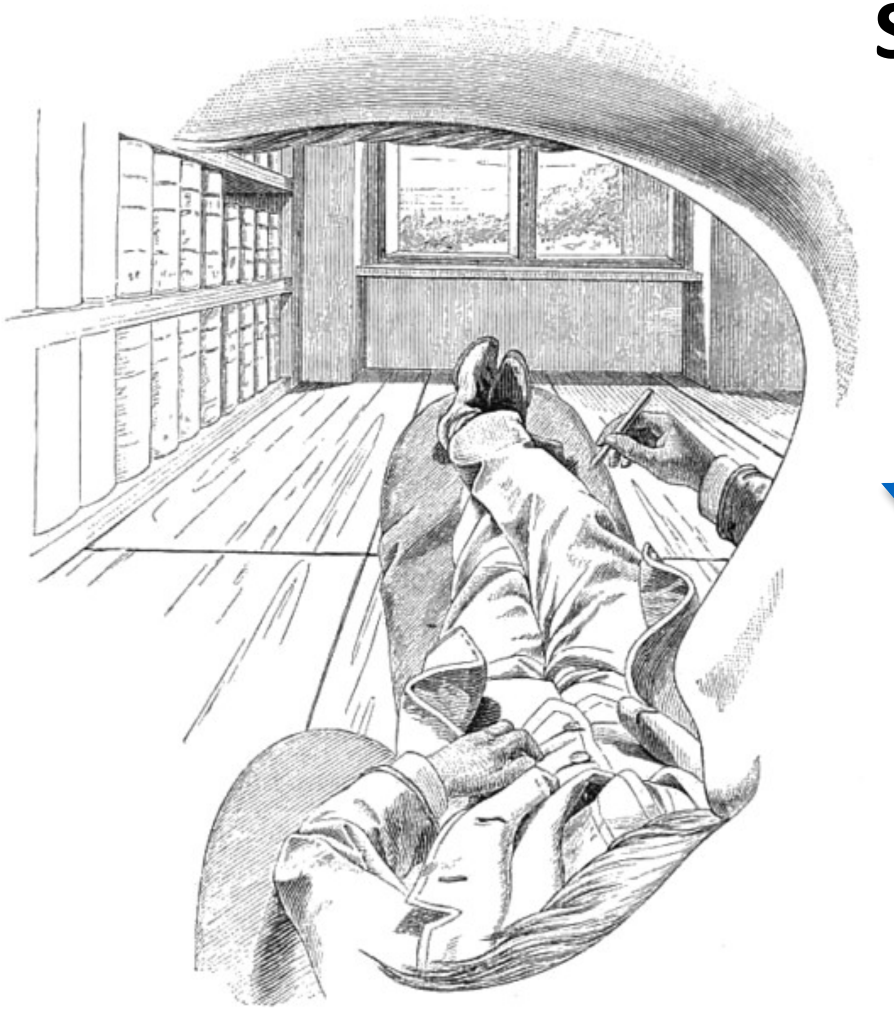


Figur 1.

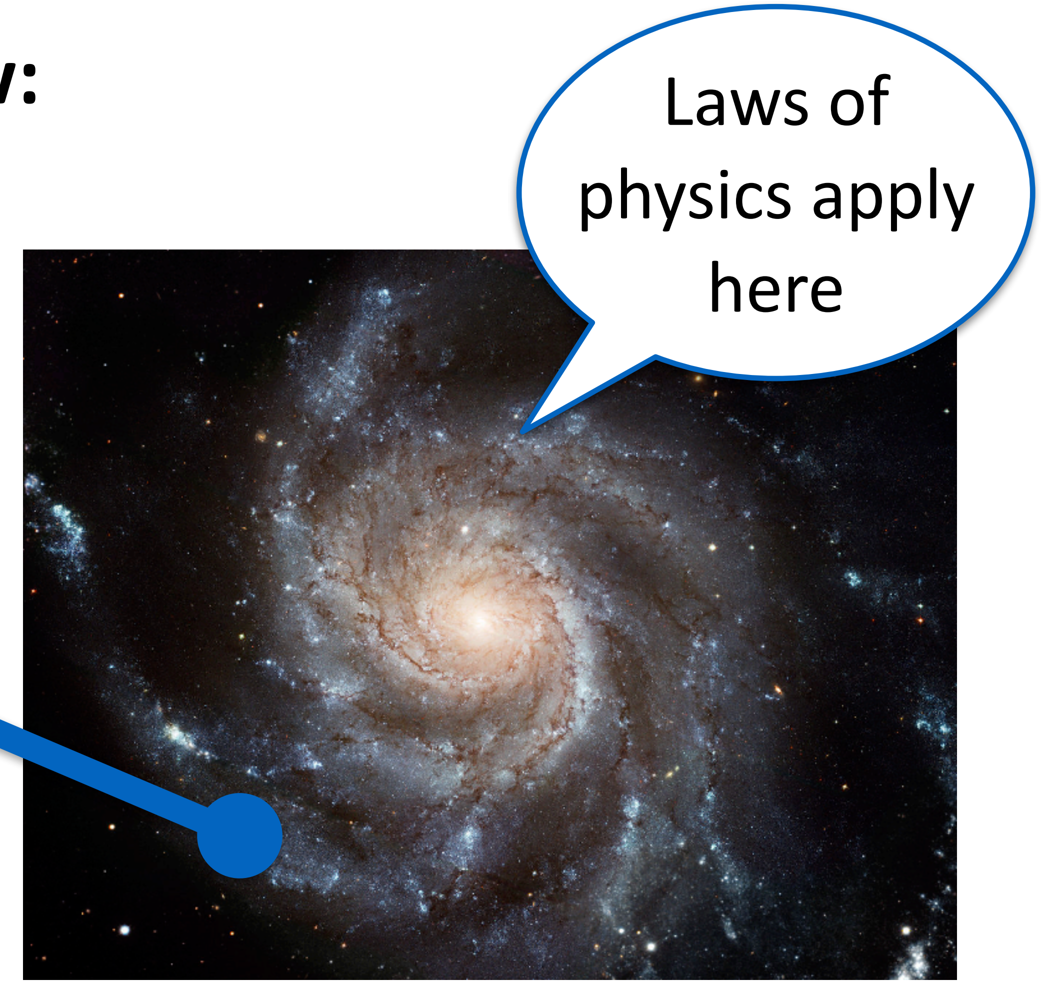


1. Conceptual puzzles

Standard view:



Figur 1.

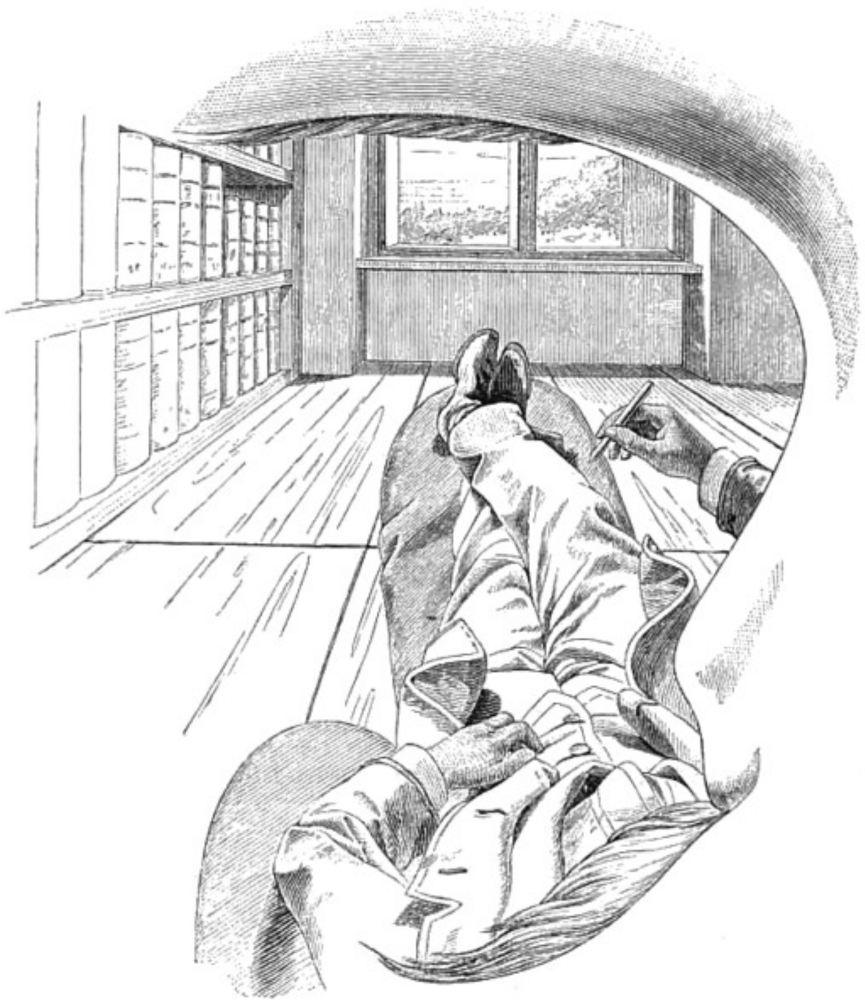


"**self pattern**" (what "I am right now", including observations and memory)

follows from

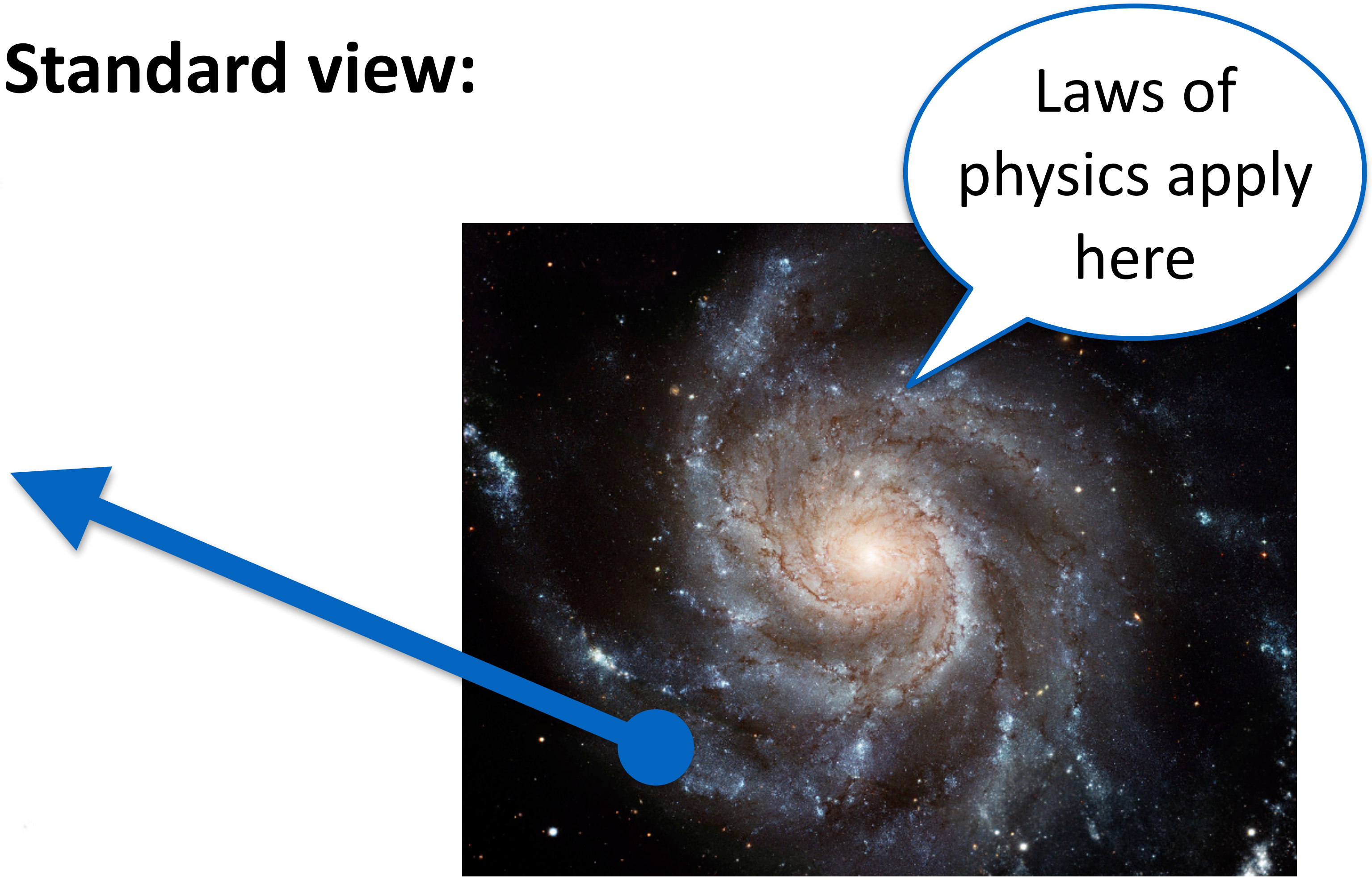
state (and evolution) of the physical world

1. Conceptual puzzles



Figur 1.

Standard view:



follows from state (and evolution) of the physical world

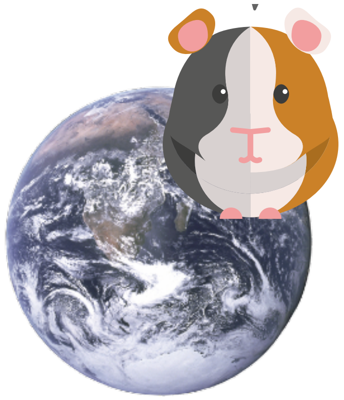
Standard methodology: to predict what happens to me next, I use physics to predict the evolution of the world, and then locate myself inside it.

Methodological inadequacy of the standard view

- **Parfit's Teletransportation Paradox**

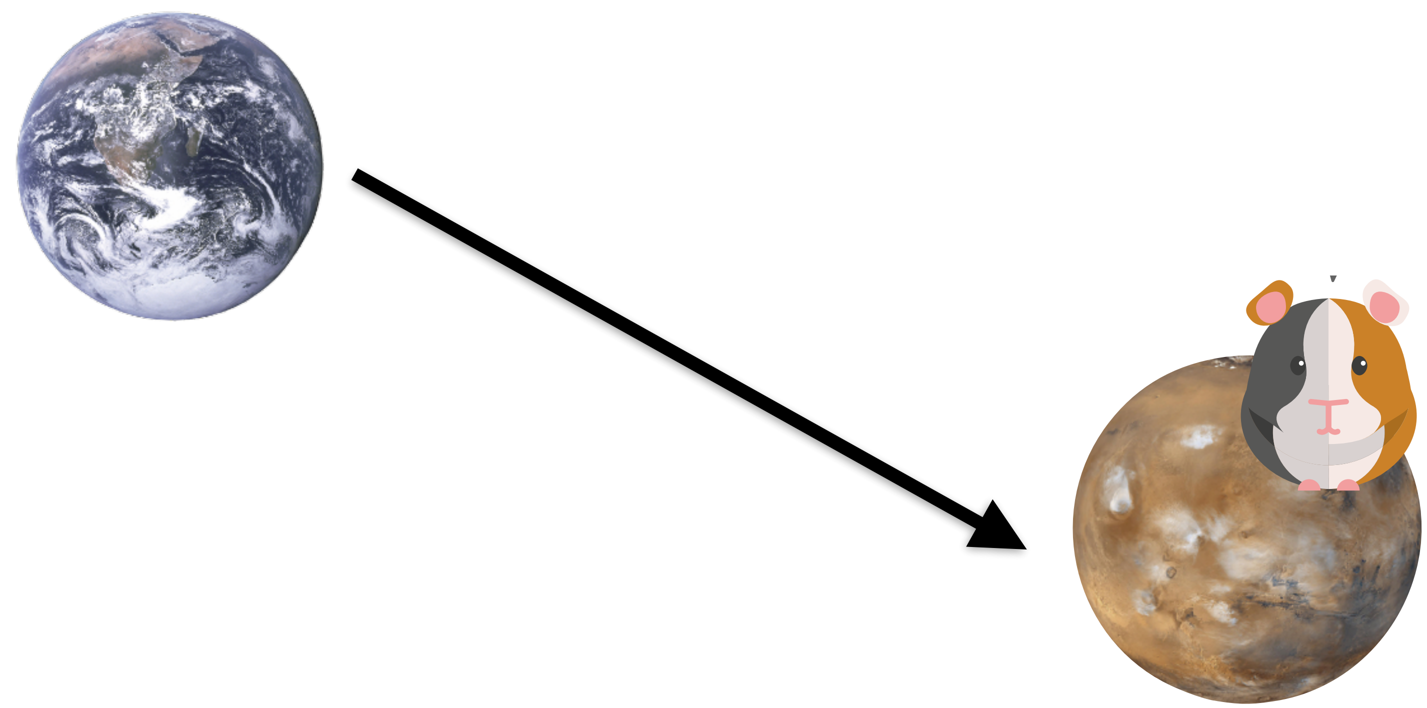
Methodological inadequacy of the standard view

- **Parfit's Teletransportation Paradox**



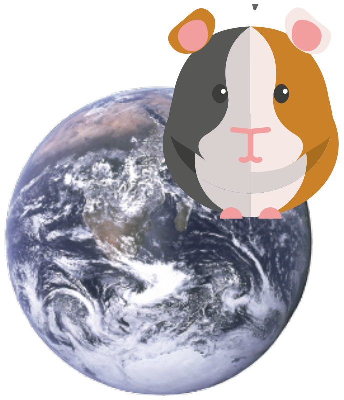
Methodological inadequacy of the standard view

- **Parfit’s Teletransportation Paradox**



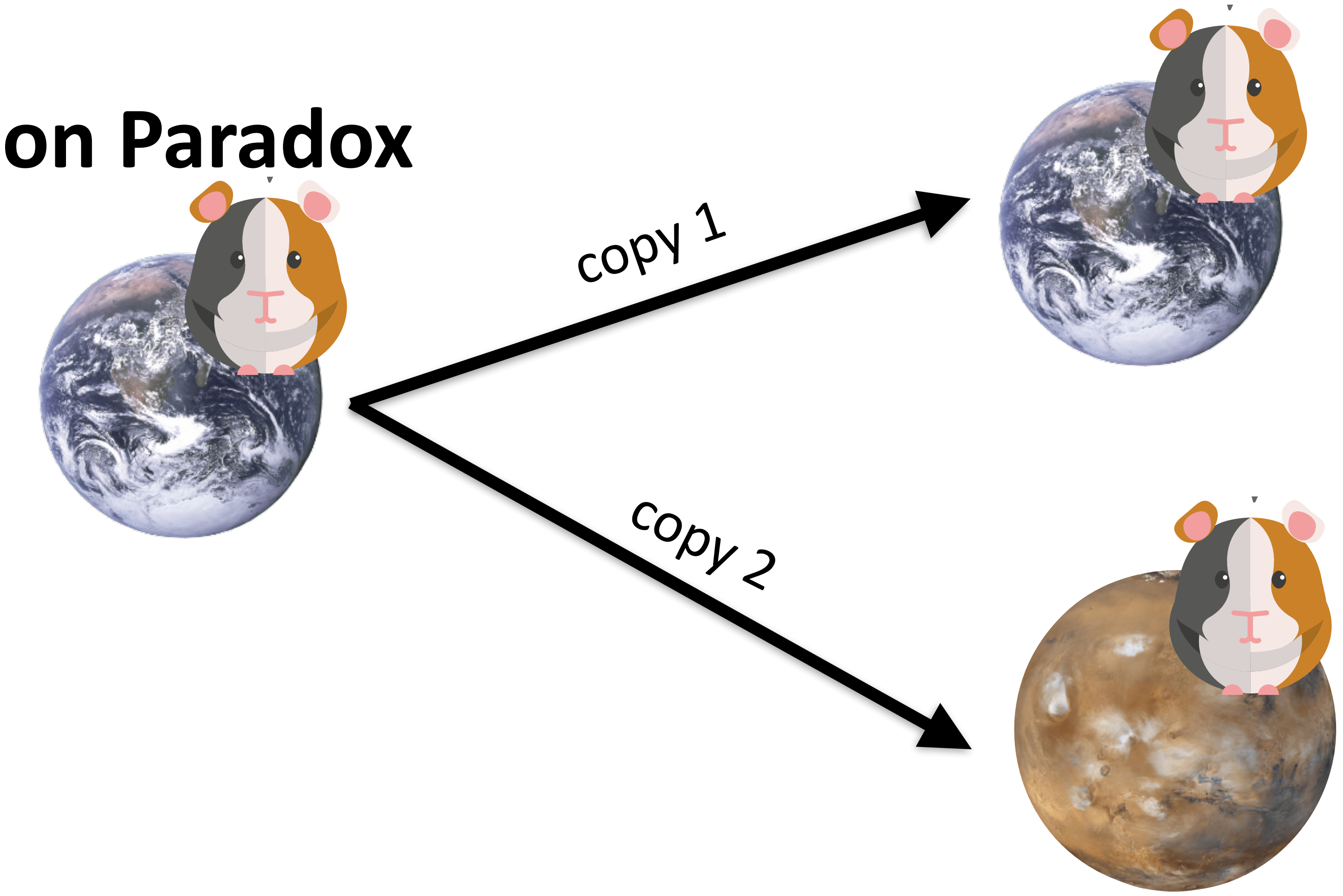
Methodological inadequacy of the standard view

- **Parfit's Teletransportation Paradox**



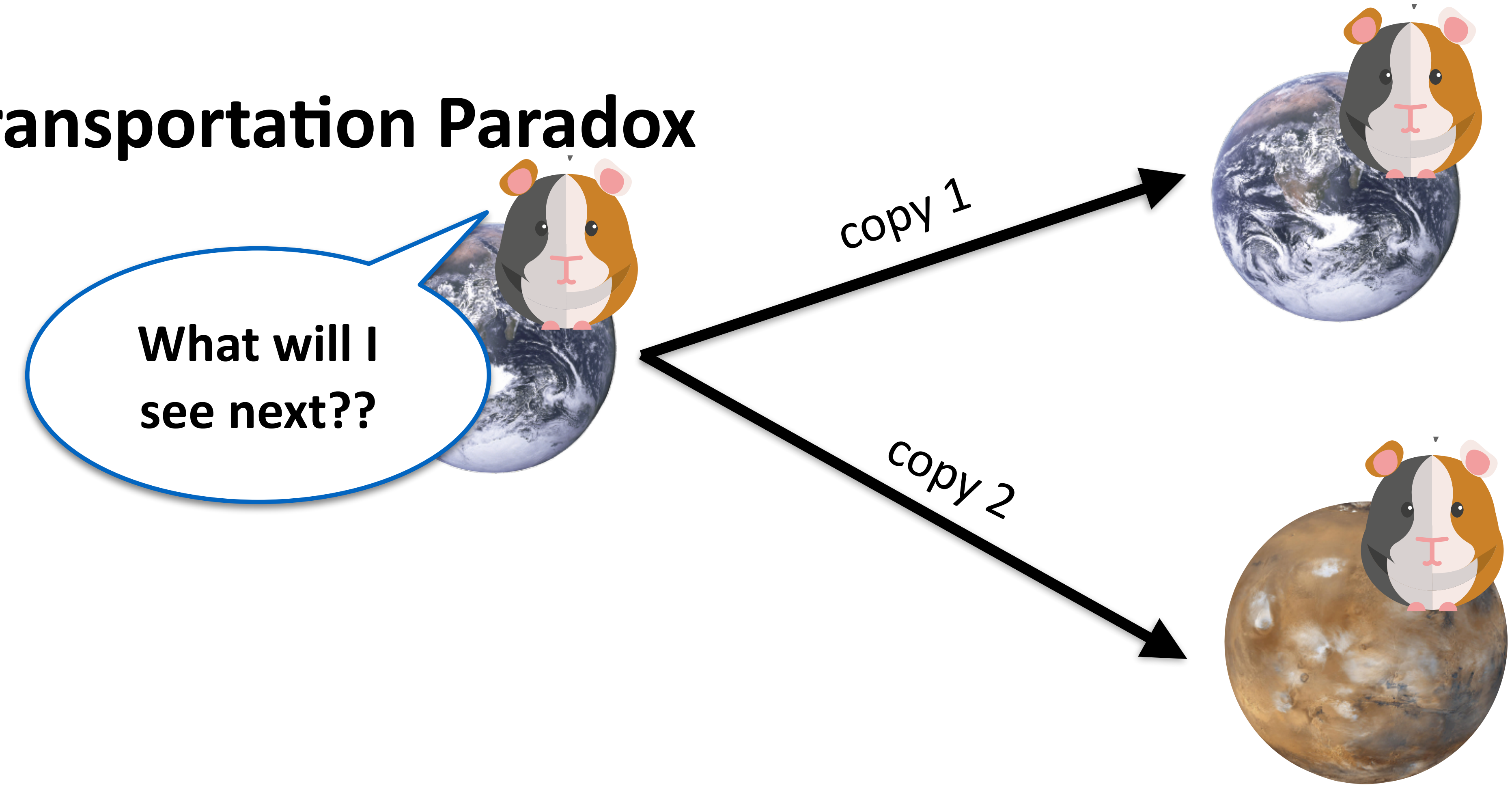
Methodological inadequacy of the standard view

- **Parfit's Teletransportation Paradox**



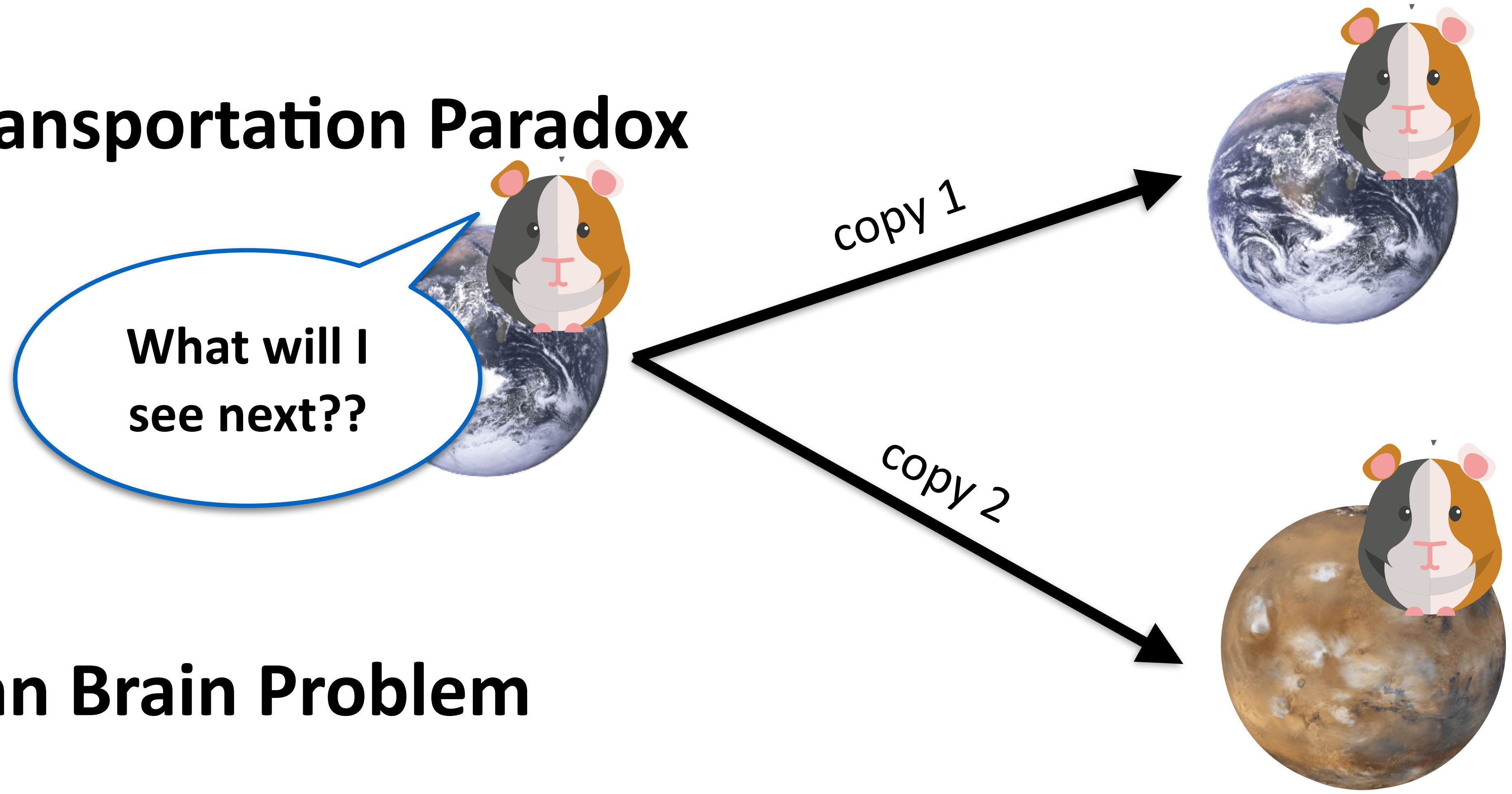
Methodological inadequacy of the standard view

- **Parfit's Teletransportation Paradox**



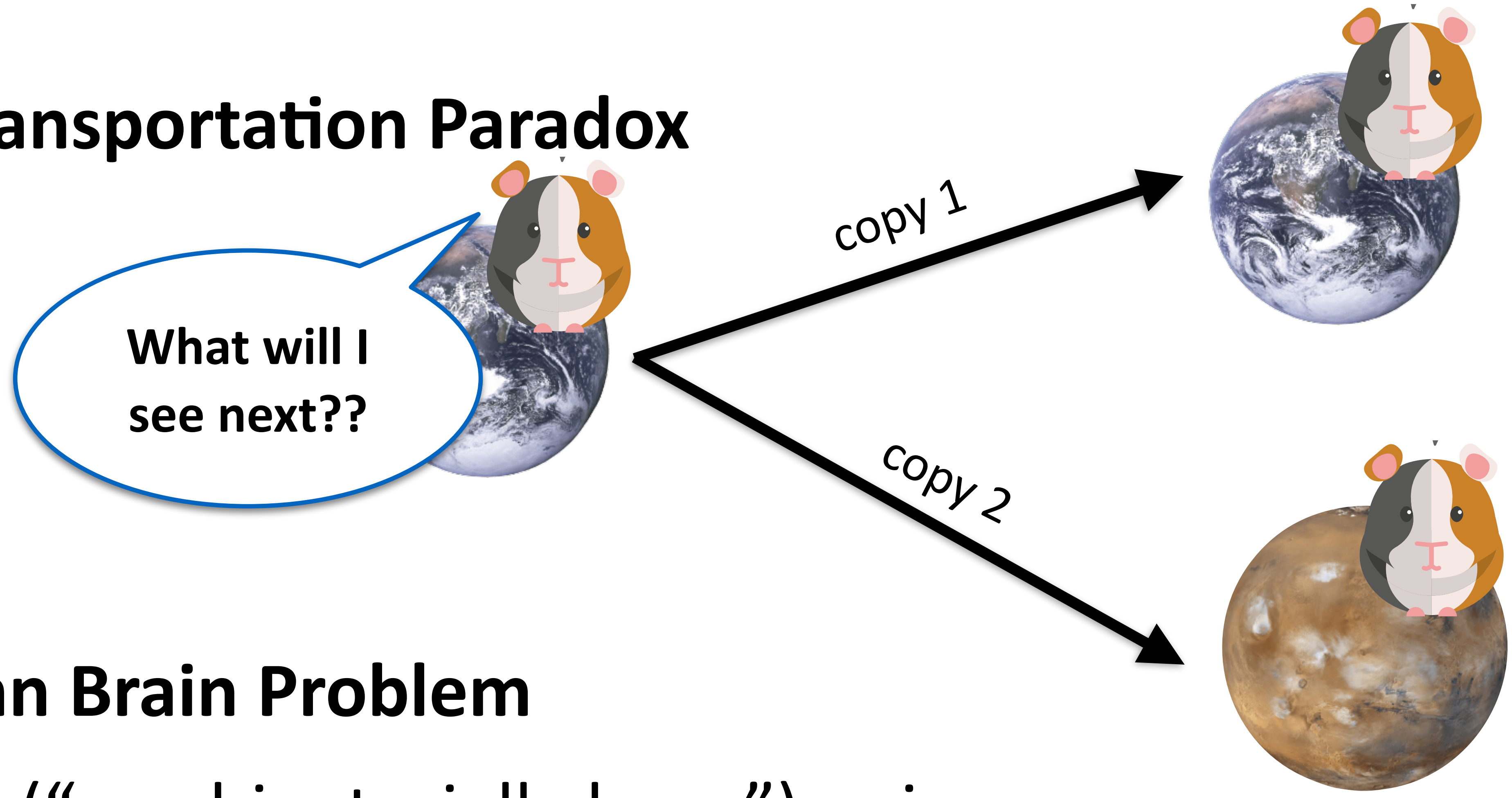
Methodological inadequacy of the standard view

- **Parfit's Teletransportation Paradox**



- **The Boltzmann Brain Problem**

- **Parfit's Teletransportation Paradox**

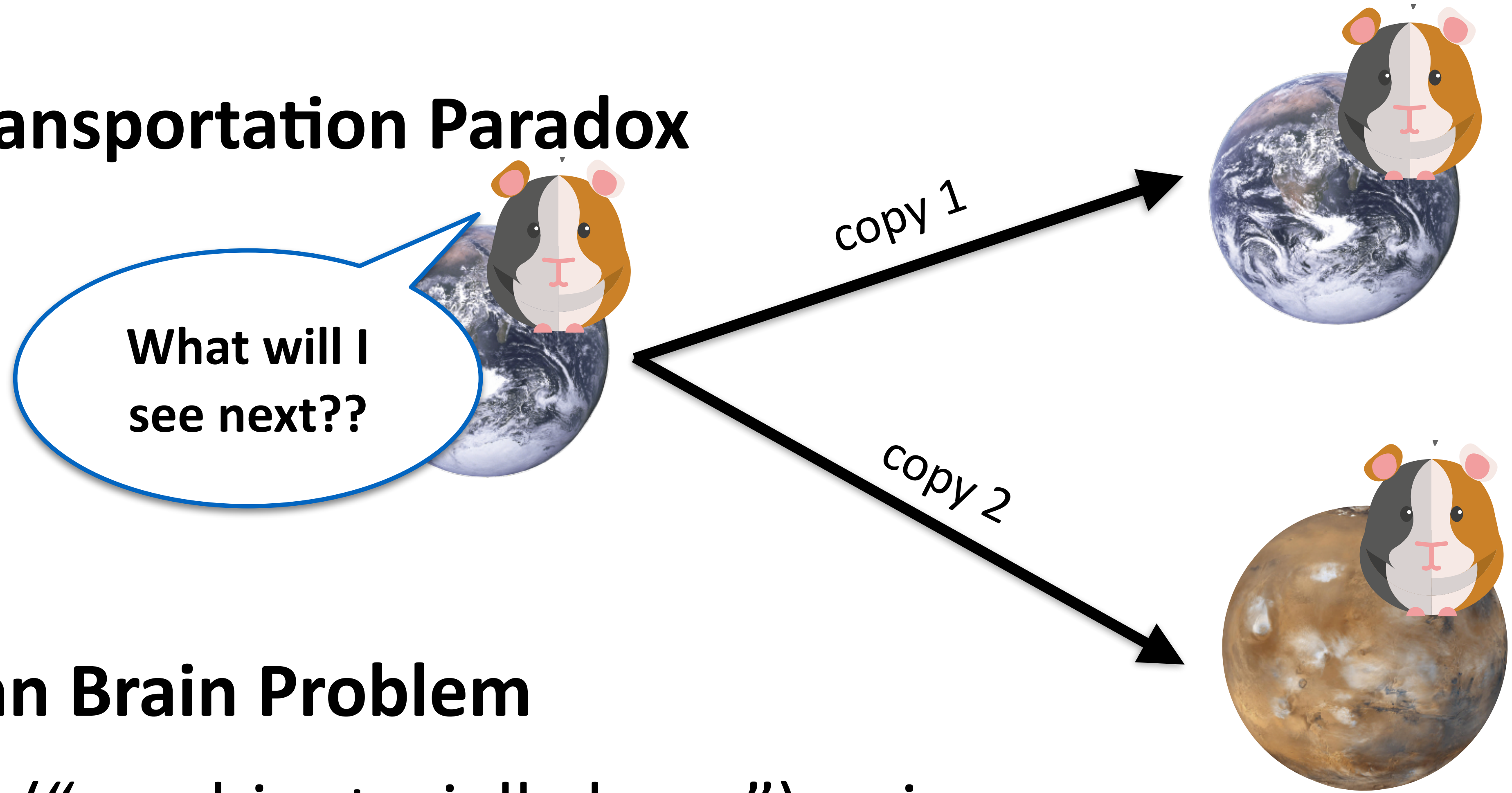


- **The Boltzmann Brain Problem**

  Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

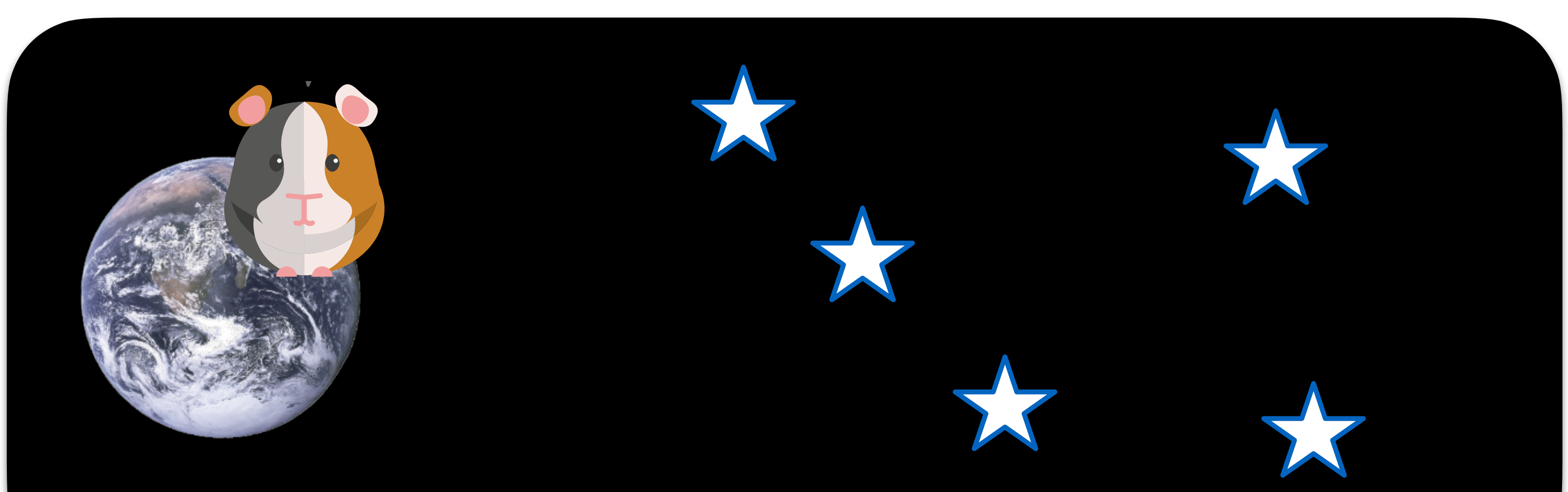
Methodological inadequacy of the standard view

- **Parfit's Teletransportation Paradox**



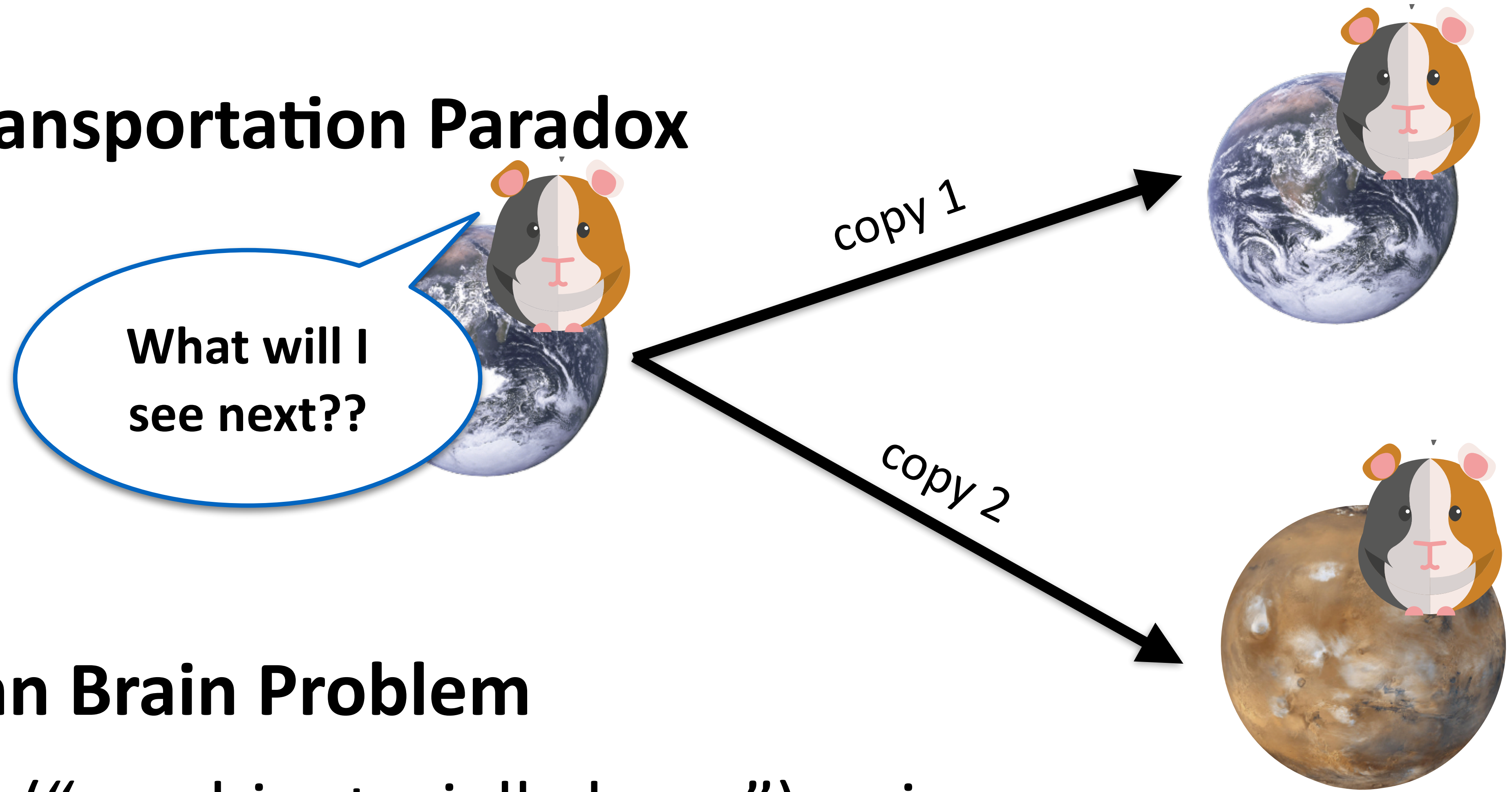
- **The Boltzmann Brain Problem**

  Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.



Methodological inadequacy of the standard view

- **Parfit’s Teletransportation Paradox**

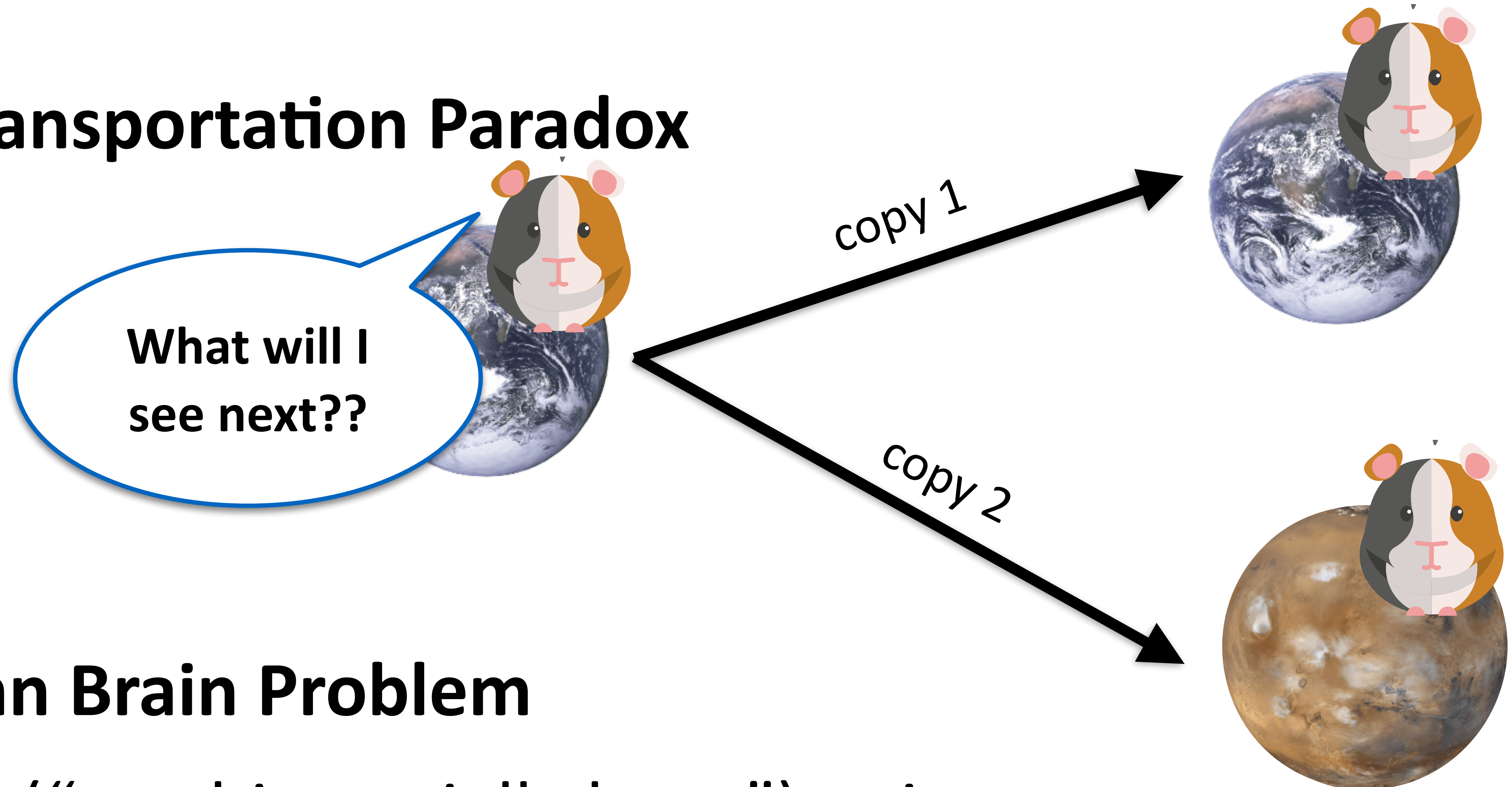


- **The Boltzmann Brain Problem**

  Assume some (“combinatorially large”) universe with a large number of “brains” with false memories fluctuating into existence.

Methodological inadequacy of the standard view

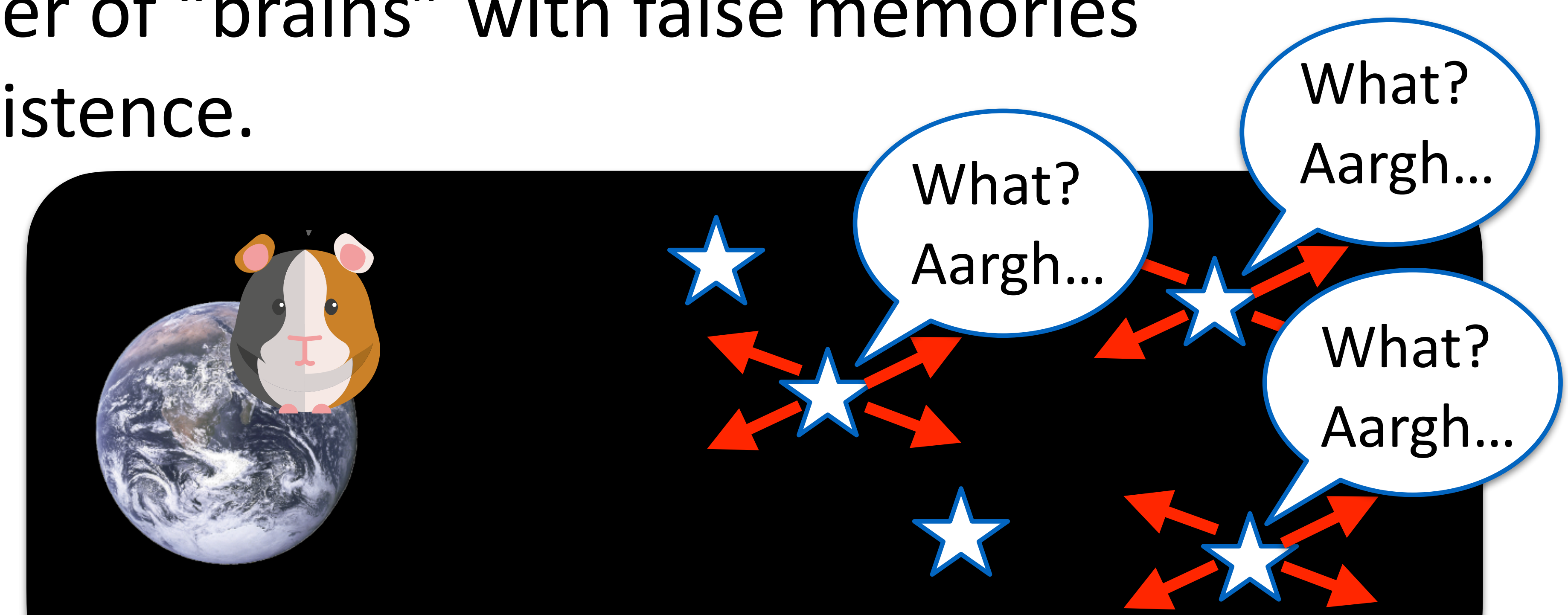
- **Parfit's Teletransportation Paradox**



- **The Boltzmann Brain Problem**

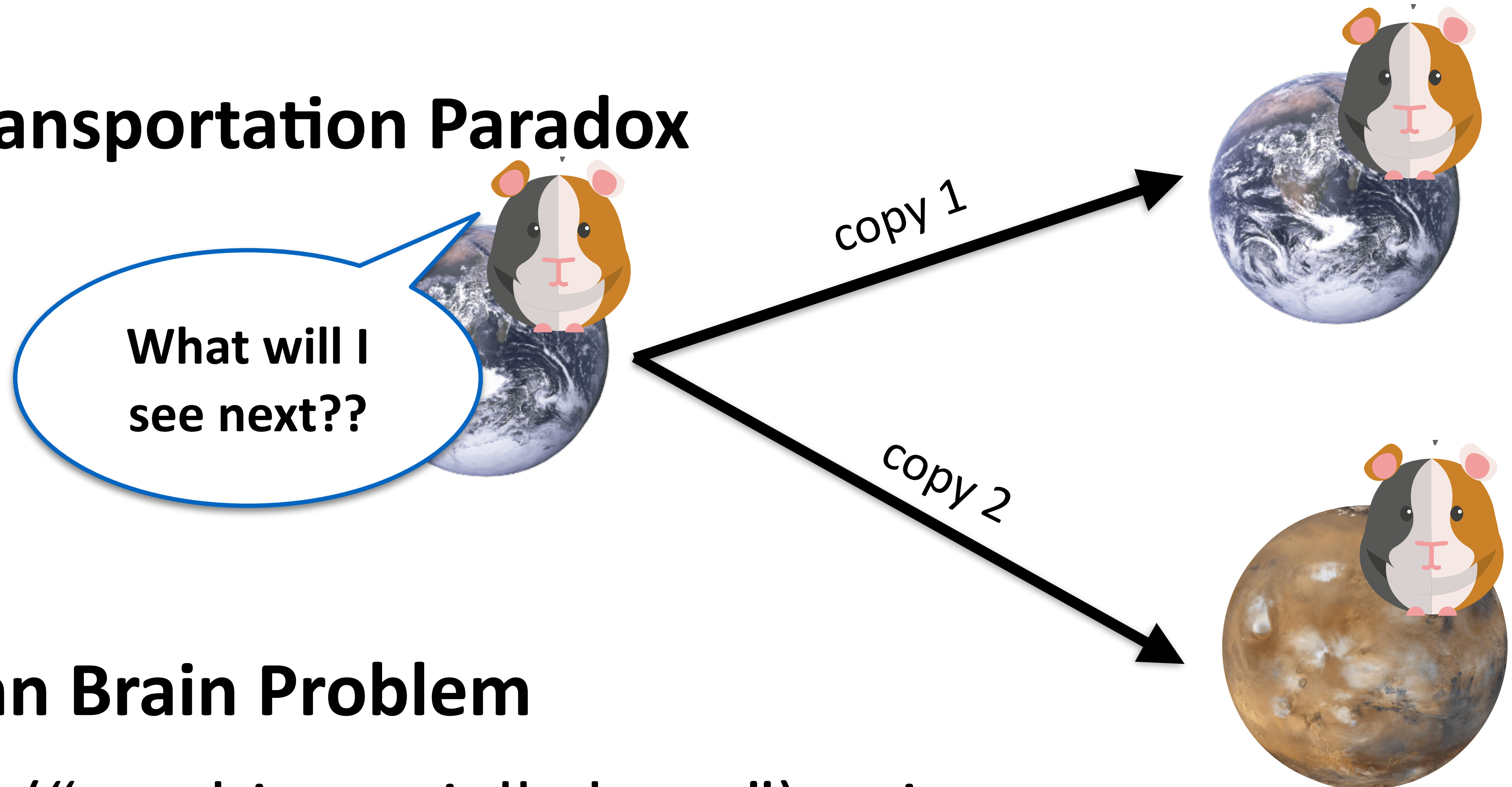
  Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

next moment



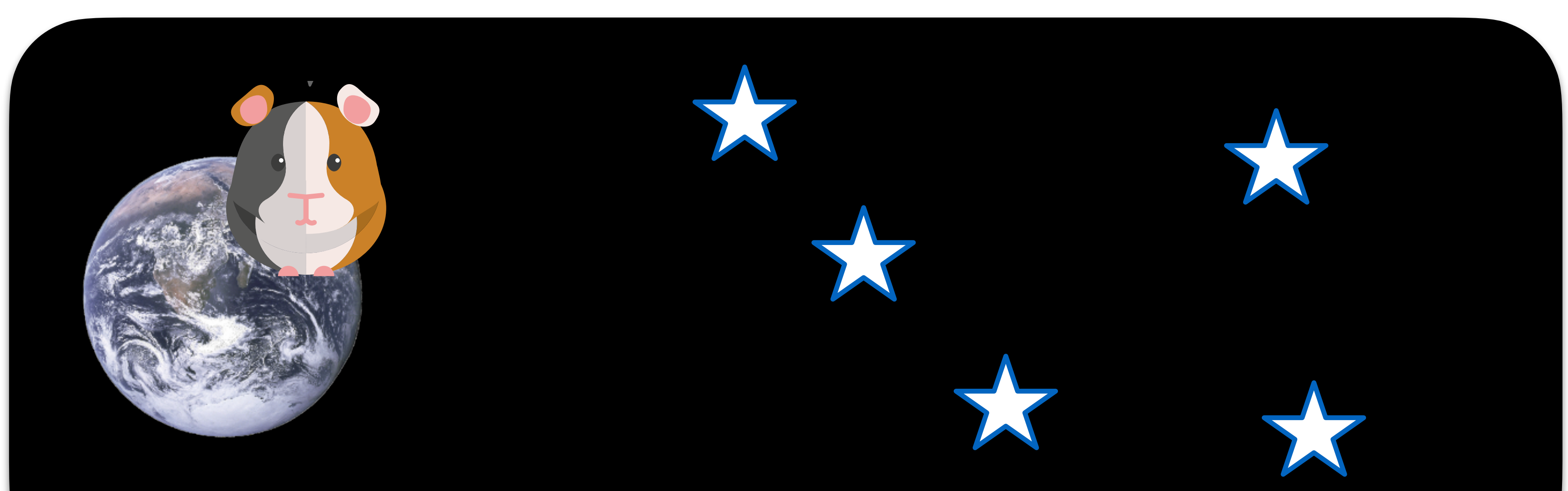
Methodological inadequacy of the standard view

- **Parfit's Teletransportation Paradox**

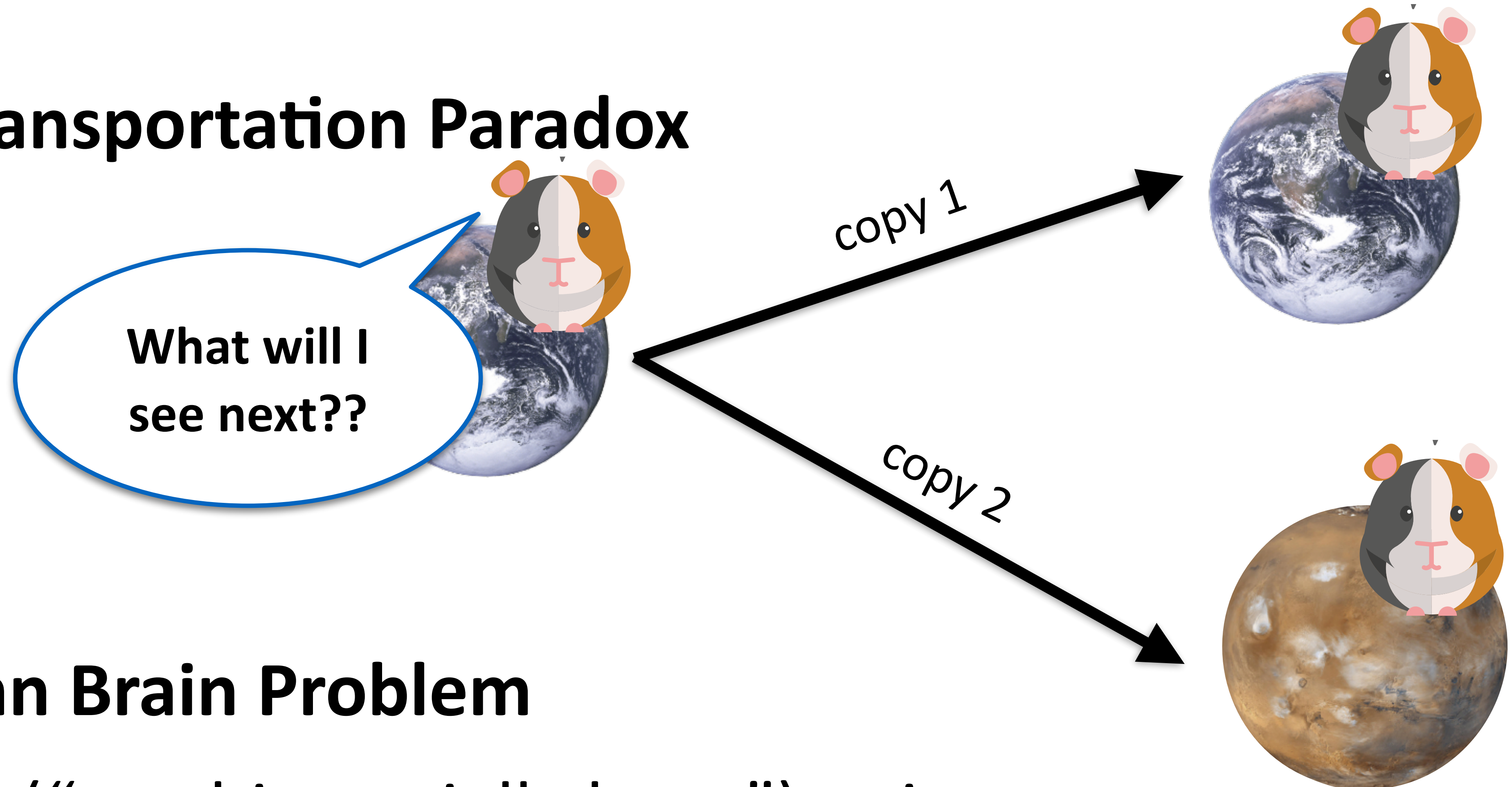


- **The Boltzmann Brain Problem**

  Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.



- **Parfit's Teletransportation Paradox**

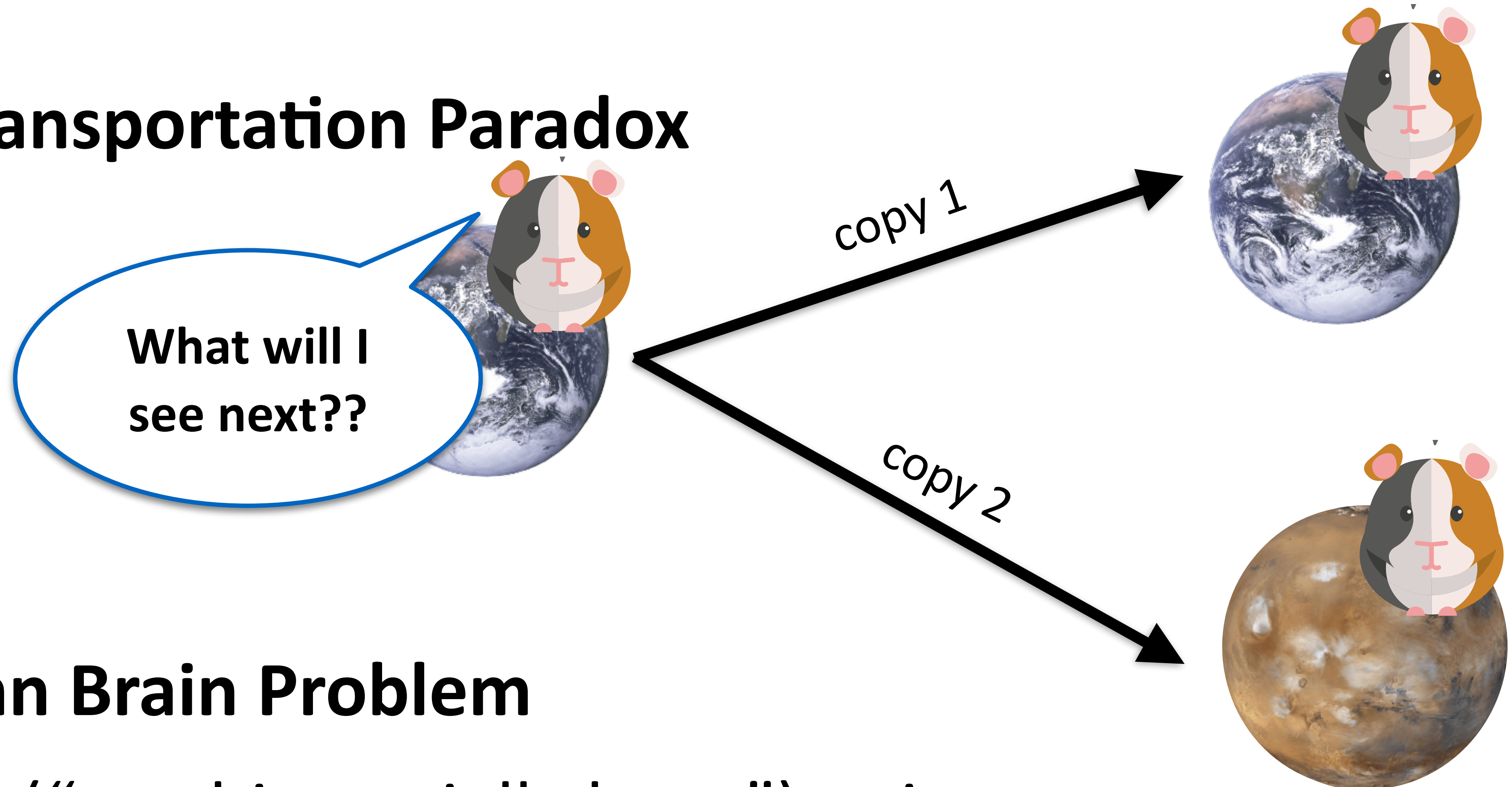


- **The Boltzmann Brain Problem**

Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

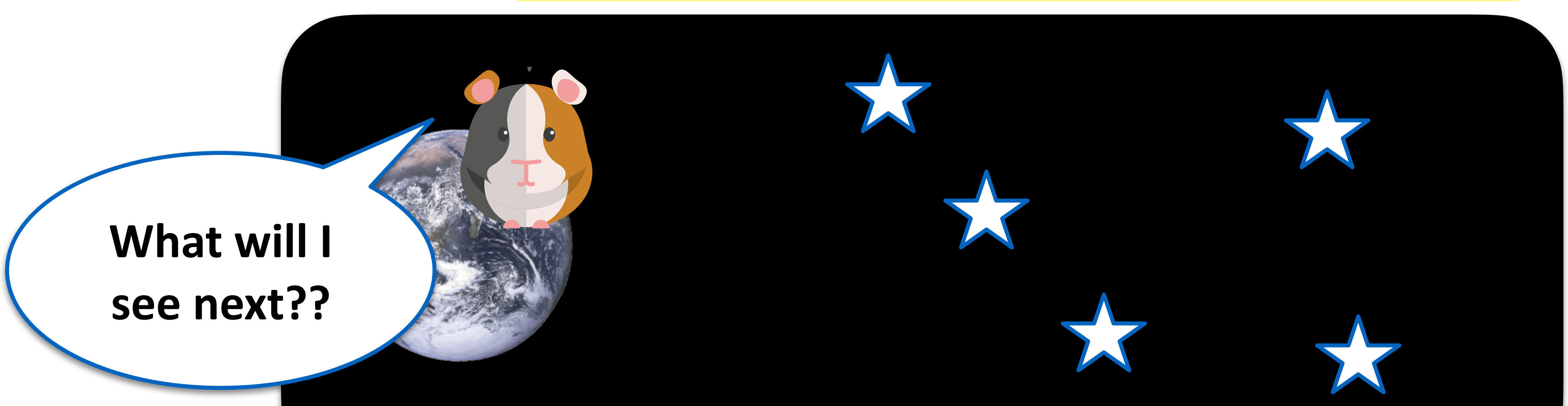
Methodological inadequacy of the standard view

- **Parfit's Teletransportation Paradox**



- **The Boltzmann Brain Problem**

  Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence. **Excludes some cosmological models?**



Methodological inadequacy of the standard view

- **Wigner's Friend**

nature physics

ARTICLES

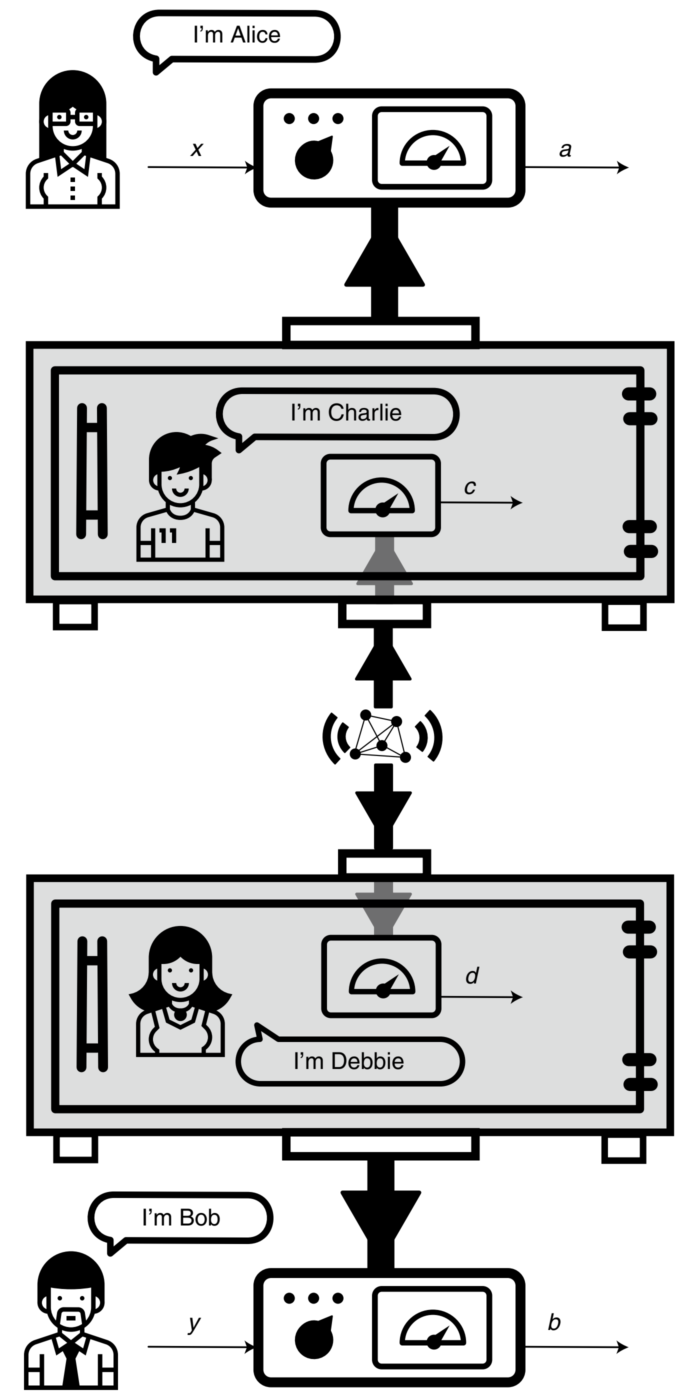
https://doi.org/10.1038/s41567-020-0990-x

Check for updates

A strong no-go theorem on the Wigner's friend paradox

Kok-Wei Bong[1,4], Aníbal Utreras-Alarcón[1,4], Farzad Ghafari[1], Yeong-Cherng Liang[2], Nora Tischler[1]✉, Eric G. Cavalcanti[3]✉, Geoff J. Pryde[1] and Howard M. Wiseman[1]

Does quantum theory apply at all scales, including that of observers? New light on this fundamental question has recently been shed through a resurgence of interest in the long-standing Wigner's friend paradox. This is a thought experiment addressing the quantum measurement problem—the difficulty of reconciling the (unitary, deterministic) evolution of isolated systems and the (non-unitary, probabilistic) state update after a measurement. Here, by building on a scenario with two separated but entangled friends introduced by Brukner, we prove that if quantum evolution is controllable on the scale of an observer, then one of 'No-Superdeterminism', 'Locality' or 'Absoluteness of Observed Events'—that every observed event exists absolutely, not relatively—must be false. We show that although the violation of Bell-type inequalities in such scenarios is not in general sufficient to demonstrate the contradiction between those three assumptions, new inequalities can be derived, in a theory-independent manner, that are violated by quantum correlations. This is demonstrated in a proof-of-principle experiment where a photon's path is deemed an observer. We discuss how this new theorem places strictly stronger constraints on physical reality than Bell's theorem.



"Absoluteness of observed events"?

Methodological inadequacy of the standard view

- **Wigner's Friend**

nature physics

ARTICLES

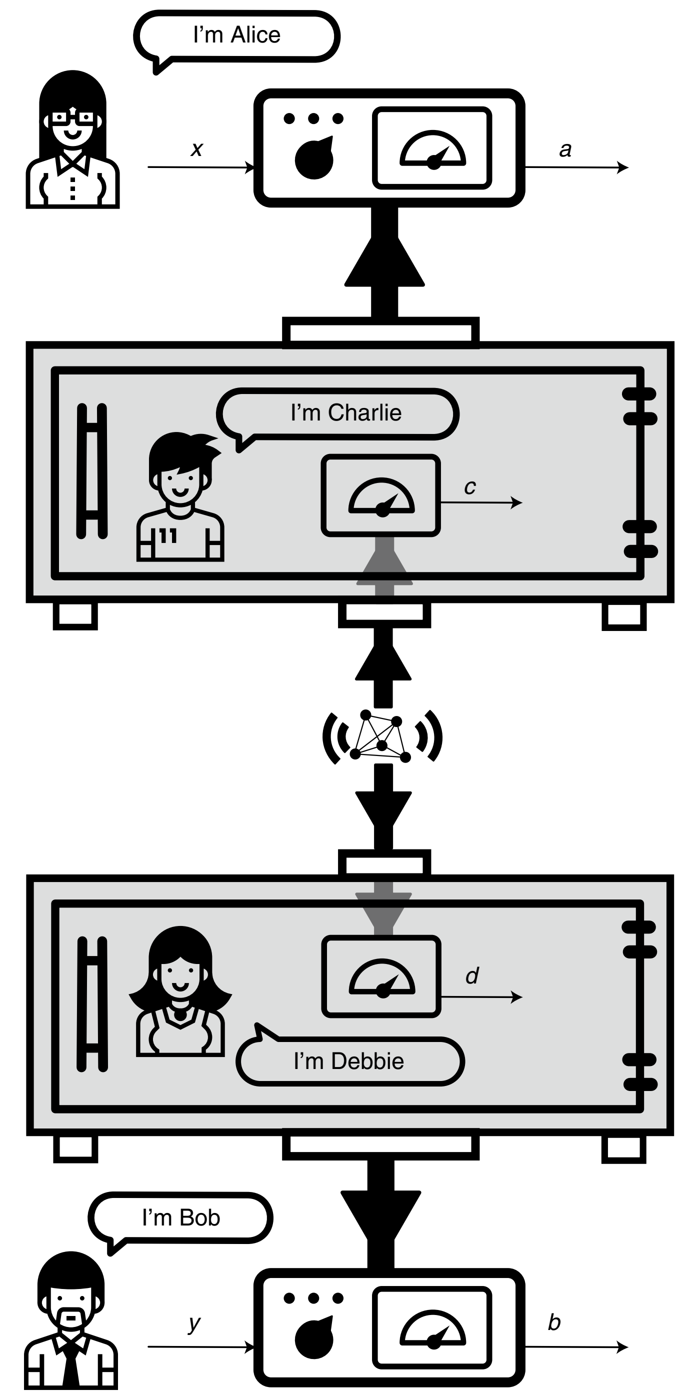
https://doi.org/10.1038/s41567-020-0990-x

Check for updates

A strong no-go theorem on the Wigner's friend paradox

Kok-Wei Bong[1,4], Aníbal Utreras-Alarcón[1,4], Farzad Ghafari[1], Yeong-Cherng Liang[2], Nora Tischler[1]✉, Eric G. Cavalcanti[3]✉, Geoff J. Pryde[1] and Howard M. Wiseman[1]

Does quantum theory apply at all scales, including that of observers? New light on this fundamental question has recently been shed through a resurgence of interest in the long-standing Wigner's friend paradox. This is a thought experiment addressing the quantum measurement problem—the difficulty of reconciling the (unitary, deterministic) evolution of isolated systems and the (non-unitary, probabilistic) state update after a measurement. Here, by building on a scenario with two separated but entangled friends introduced by Brukner, we prove that if quantum evolution is controllable on the scale of an observer, then one of 'No-Superdeterminism', 'Locality' or 'Absoluteness of Observed Events'—that every observed event exists absolutely, not relatively—must be false. We show that although the violation of Bell-type inequalities in such scenarios is not in general sufficient to demonstrate the contradiction between those three assumptions, new inequalities can be derived, in a theory-independent manner, that are violated by quantum correlations. This is demonstrated in a proof-of-principle experiment where a photon's path is deemed an observer. We discuss how this new theorem places strictly stronger constraints on physical reality than Bell's theorem.



"Absoluteness of observed events"?

In a quantum world, it is unclear how to use the "standard methodology" without running into paradoxes.

“What will I see next?”

- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

“What will I see next?”

- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

exotic regime

“What will I see next?”

- **Simulating agents on a computer**
- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

exotic regime

“What will I see next?”

- **Simulating agents on a computer**
- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

exotic regime

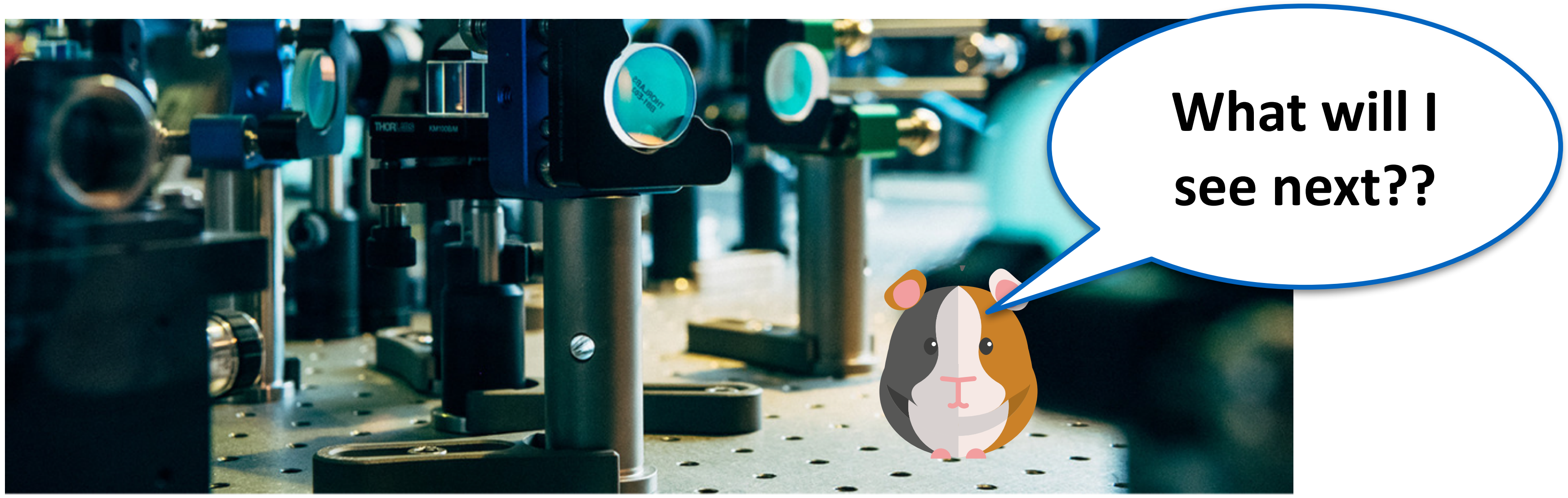
- **Laboratory experiments**

“What will I see next?”

- **Simulating agents on a computer**
- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

exotic regime

- **Laboratory experiments**



"What will I see next?"

- **Simulating agents on a computer**
- **Parfit's Teletransportation Paradox**
- **Wigner's Friend**
- **The Boltzmann Brain Problem**

exotic regime

- **Laboratory experiments**

“What will I see next?”

- **Simulating agents on a computer**
- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

exotic regime

- **Laboratory experiments**
- **Astronomical observations**

“What will I see next?”

- **Simulating agents on a computer**
- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

exotic regime

- **Laboratory experiments**
- **Astronomical observations**
- **…**

empirical regime

“What will I see next?”

- **Simulating agents on a computer**
- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

exotic regime

- **Laboratory experiments**
- **Astronomical observations**
- **…**

physics applies!

empirical regime

"What will I see next?"

- Simulating agents on a computer
- Parfit's Teletransportation Paradox
- Wigner's Friend
- The Boltzmann Brain Problem

philosophy of mind?

exotic regime

- Laboratory experiments
- Astronomical observations
- …

physics applies!

empirical regime

"What will I see next?"

- **Simulating agents on a computer**
- **Parfit's Teletransportation Paradox**
- **Wigner's Friend**
- **The Boltzmann Brain Problem**

But this appeared in physics!

philosophy of mind?

exotic regime

- **Laboratory experiments**
- **Astronomical observations**
- **…**

physics applies!

empirical regime

“What will I see next?”

- **Simulating agents on a computer**
- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

philosophy of mind?

exotic regime

- **Laboratory experiments**
- **Astronomical observations**
- **…**

physics applies!

empirical regime

"What will I see next?"

- **Simulating agents on a computer**
- **Parfit's Teletransportation Paradox**
- **Wigner's Friend**
- **The Boltzmann Brain Problem**

philosophy of mind?

exotic regime

- **Laboratory experiments**
- **Astronomical observations**
- **…**

physics applies!

empirical regime

“What will I see next?”

Wait a minute… isn’t physics concerned with a **different kind of question?**

“What will I see next?”

Wait a minute… isn’t physics concerned with a **different kind of question?**

“What is the world like?” instead of “what will I see”…

“What will I see next?”

Wait a minute… isn’t physics concerned with a **different kind of question?**

“What is the world like?” instead of “what will I see”…

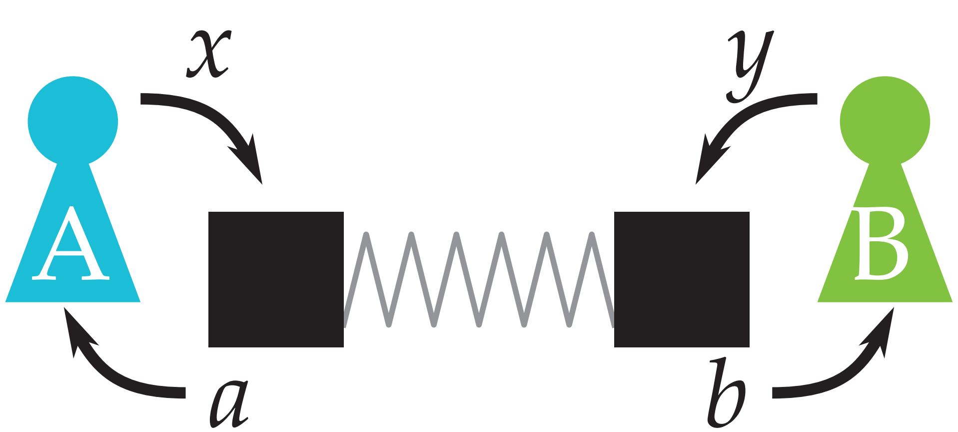
Quantum theory: The formalism tells us **the probabilities of the outcomes we will see**, given our choice of measurement.

“What will I see next?”

Wait a minute… isn’t physics concerned with a **different kind of question?**

“What is the world like?” instead of “what will I see”…

Quantum theory: The formalism tells us **the probabilities of the outcomes we will see**, given our choice of measurement.



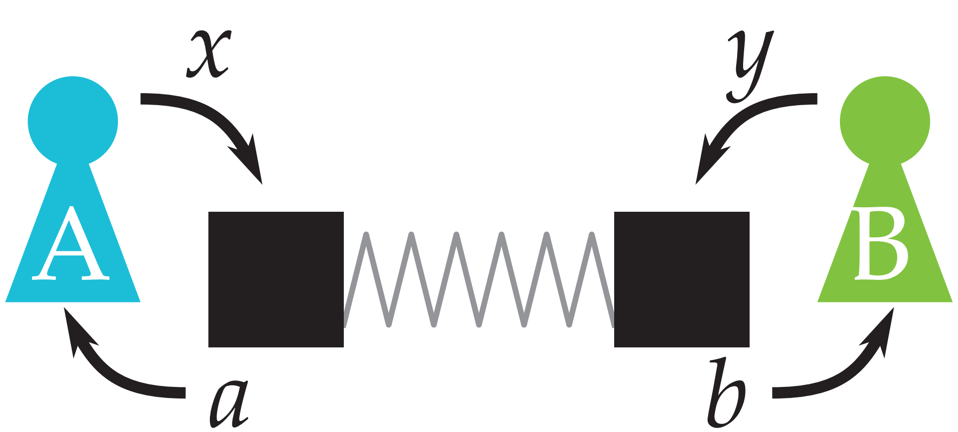
Bell’s Theorem: It is even **inconsistent** to assume that measurements always tell us what the world has been like unless we give up locality.

“What will I see next?”

Wait a minute… isn’t physics concerned with a **different kind of question?**

“What is the world like?” instead of “what will I see”…

Quantum theory: The formalism tells us **the probabilities of the outcomes we will see**, given our choice of measurement.



Bell’s Theorem: It is even **inconsistent** to assume that measurements always tell us what the world has been like unless we give up locality.

Independent motivation to consider “what will I see next?”
a more fruitful / natural question to ask than “what is the case?”

“What will I see next?”

- **Simulating agents on a computer**
- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

philosophy of mind?

exotic regime

- **Laboratory experiments**
- **Astronomical observations**
- **…**

physics applies!

empirical regime

Wanted: a **universal** answer to “what will I see next?”

- **Simulating agents on a computer**
- **Parfit’s Teletransportation Paradox**
- **Wigner’s Friend**
- **The Boltzmann Brain Problem**

exotic regime

A unified approach

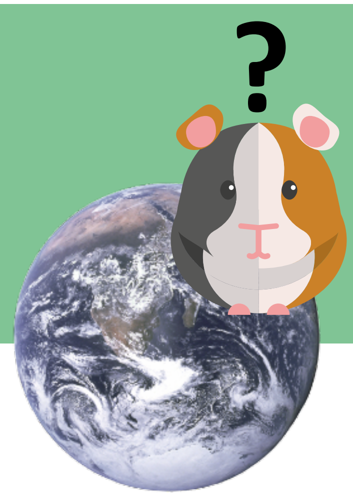
- **Laboratory experiments**
- **Astronomical observations**
- **…**

empirical regime

Outline

1. **Conceptual puzzles**

… that challenge the standard view.



2. **Sketch of an idealist (toy) theory**

… "self" fundamental, external world emergent.

3. **Objective reality as a emergent approximation**

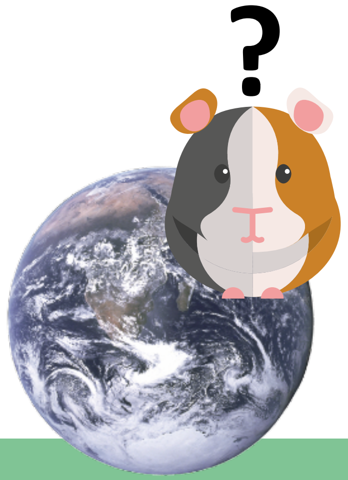
… probabilistic zombies, and other surprises.

4. **Example: dissolution of the Boltzmann brain problem**

Outline

1. **Conceptual puzzles**

… that challenge the standard view.



2. **Sketch of an idealist (toy) theory**

… "self" fundamental, external world emergent.

3. **Objective reality as a emergent approximation**

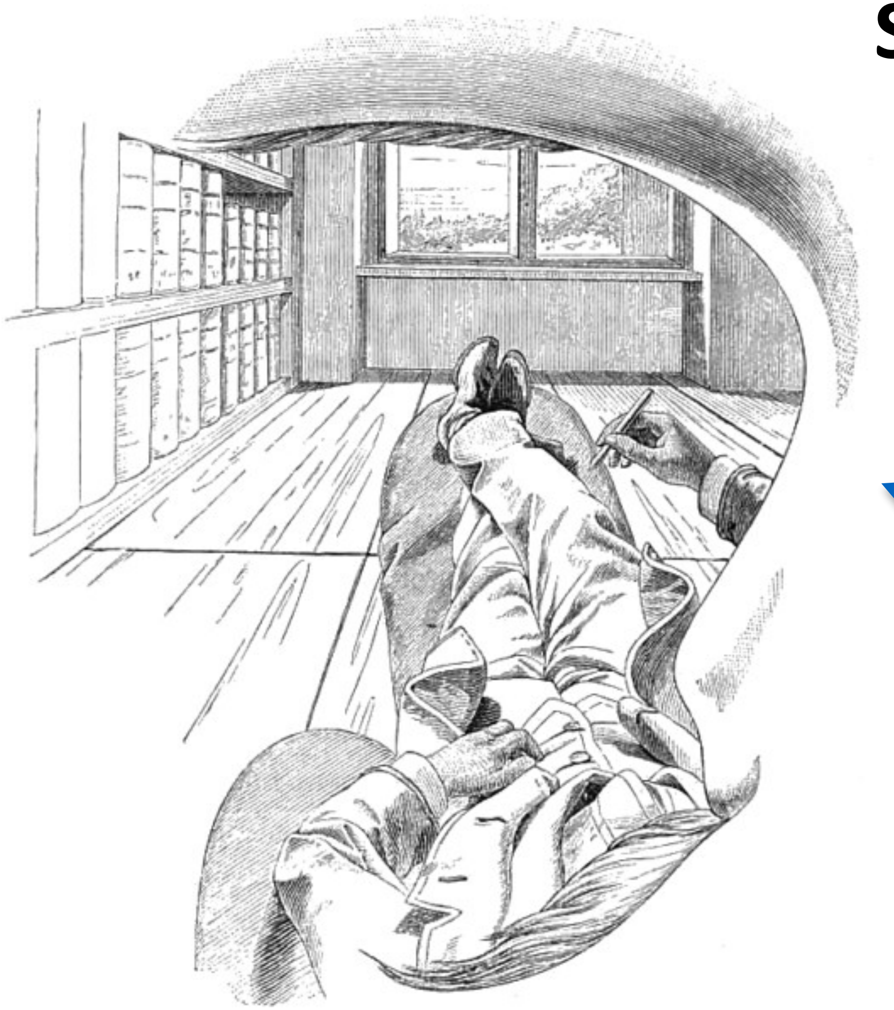
… probabilistic zombies, and other surprises.

4. **Example: dissolution of the Boltzmann brain problem**

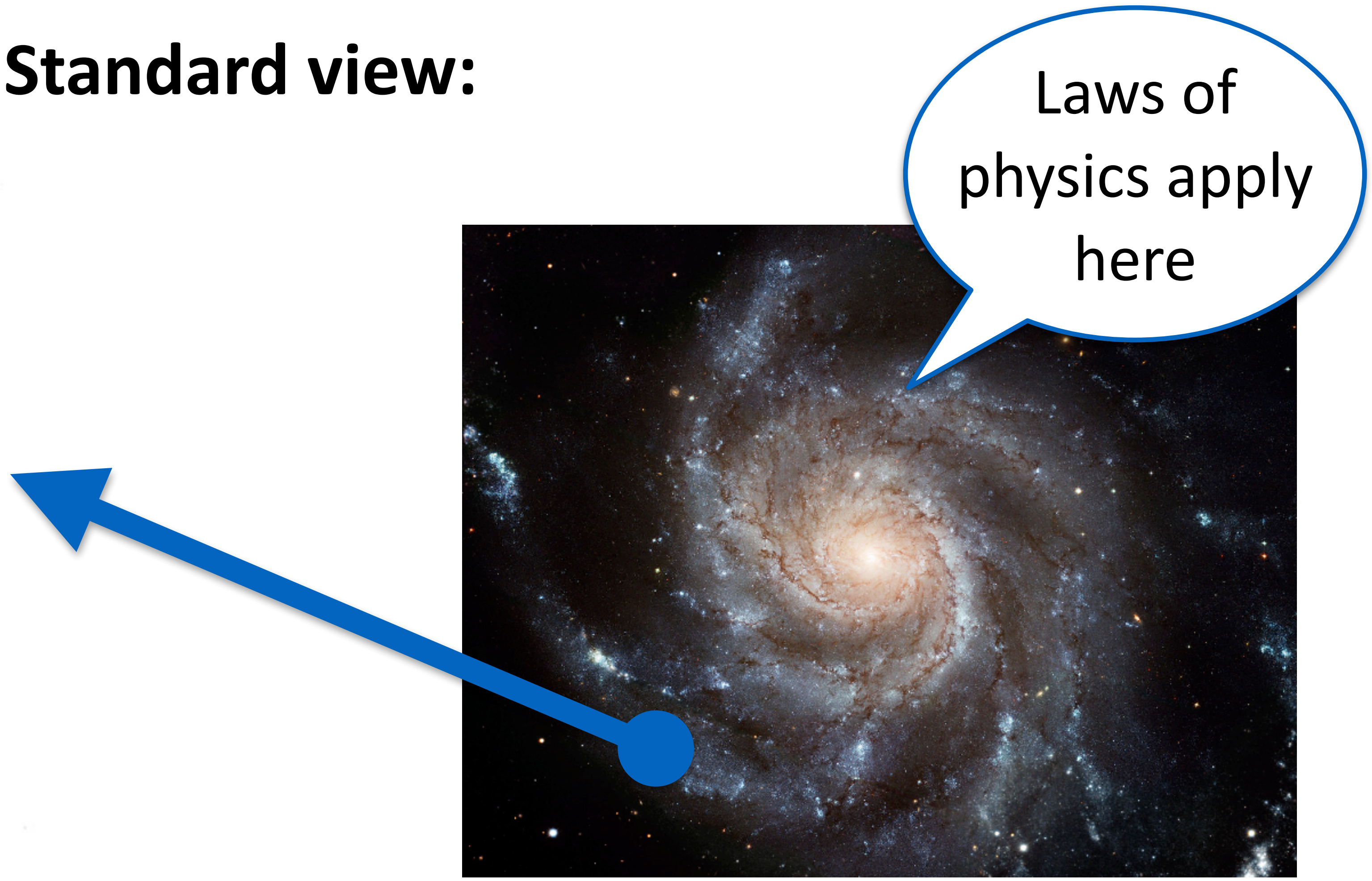
A **unified approach** by reversing the standard view

A **unified approach** by reversing the standard view

Standard view:

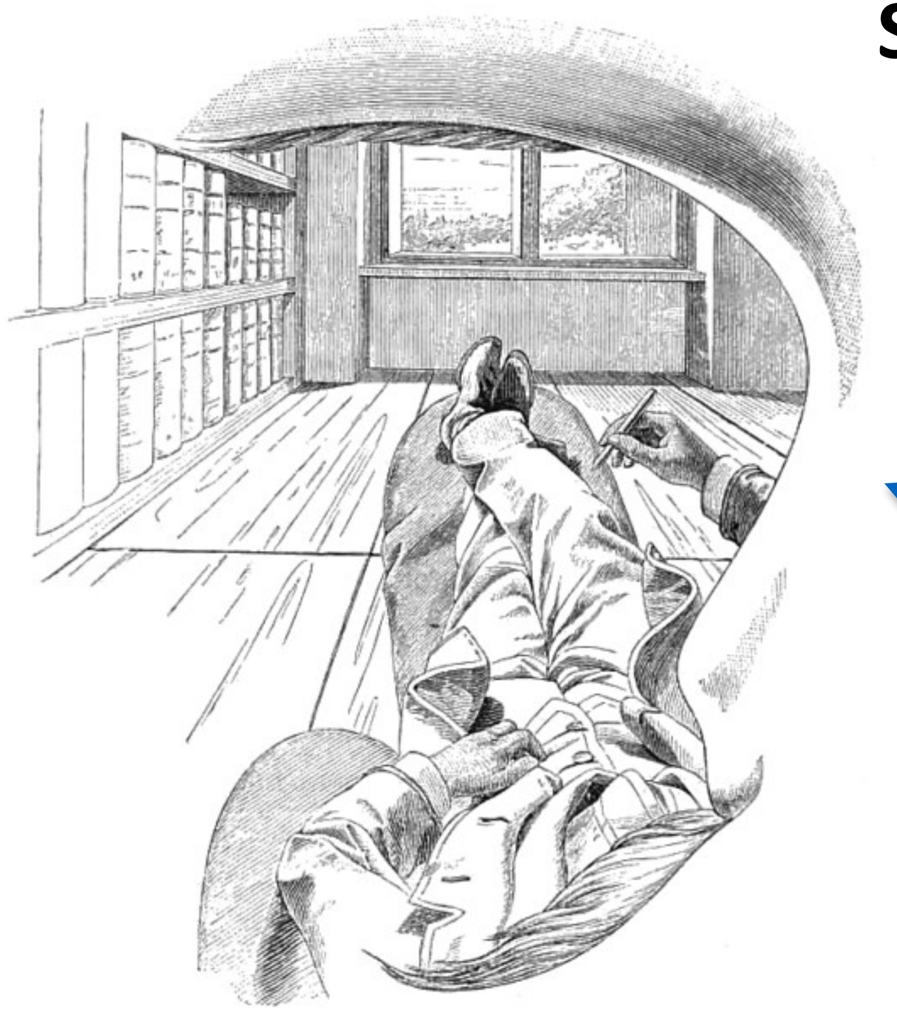


Figur 1.

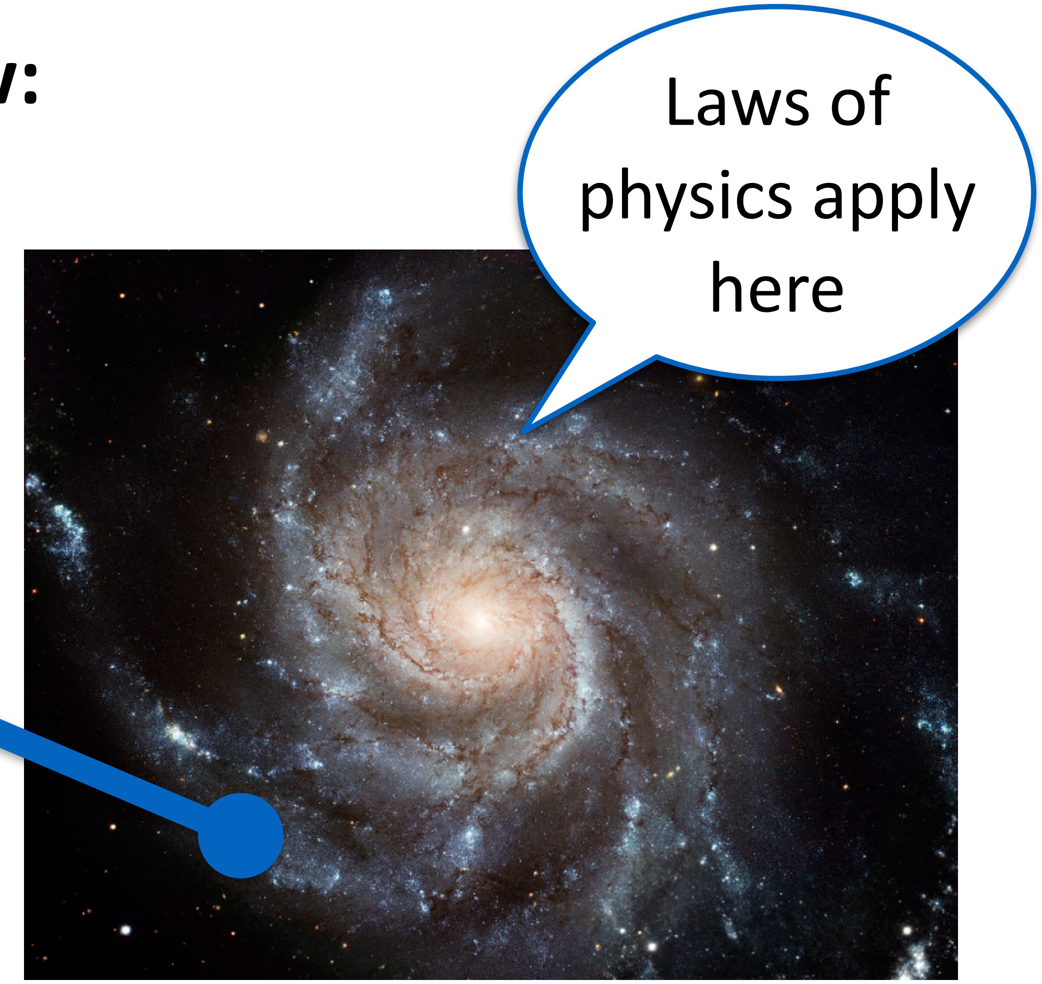


A **unified approach** by reversing the standard view

Standard view:



Figur 1.



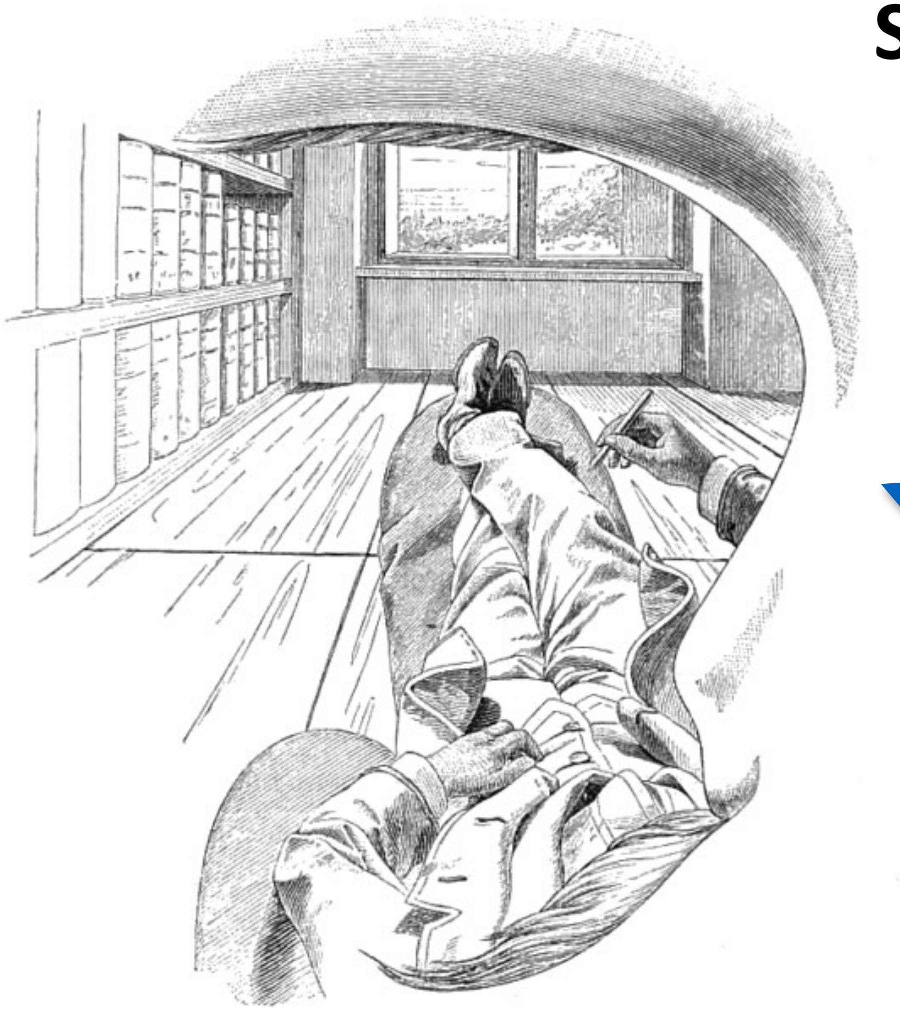
"**self pattern"** (what "I am right now", including observations and memory)

follows from

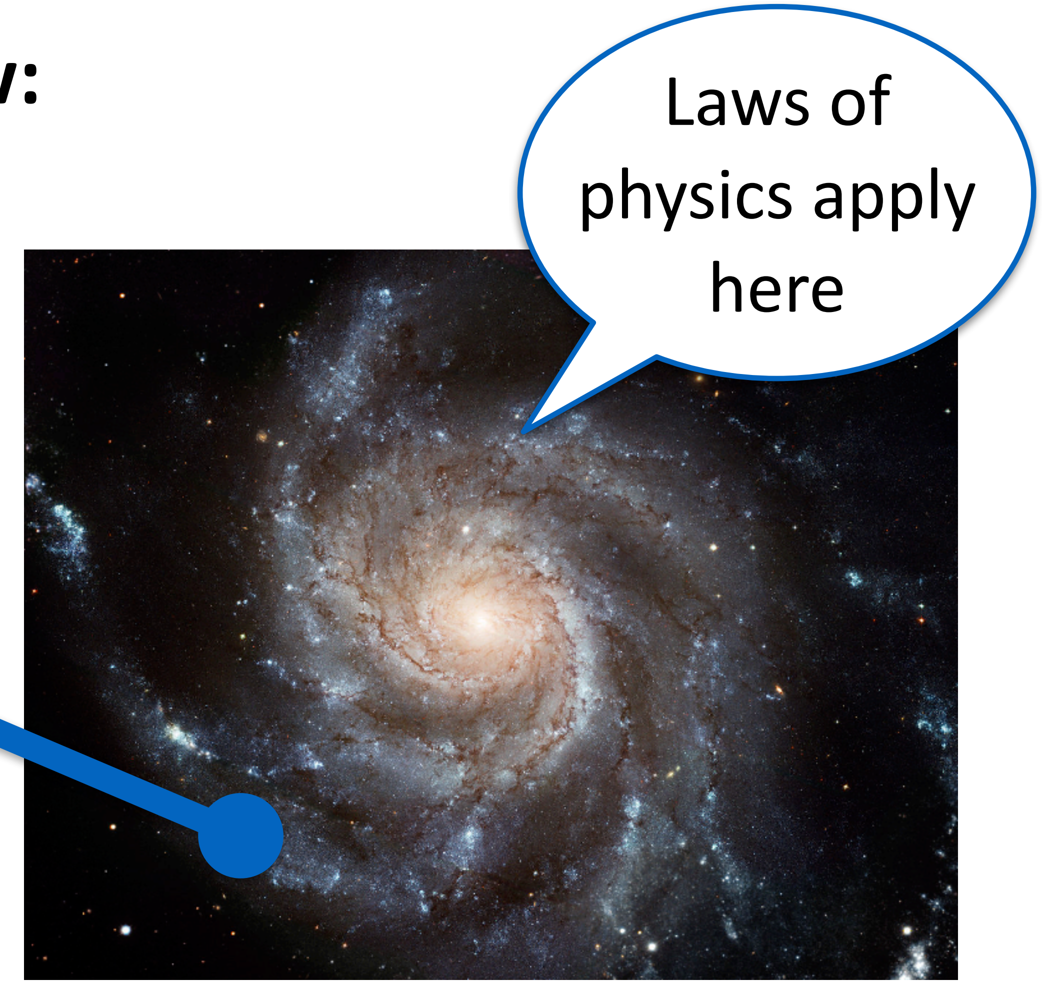
state (and evolution) of the physical world

A **unified approach** by reversing the standard view

Standard view:



Figur 1.



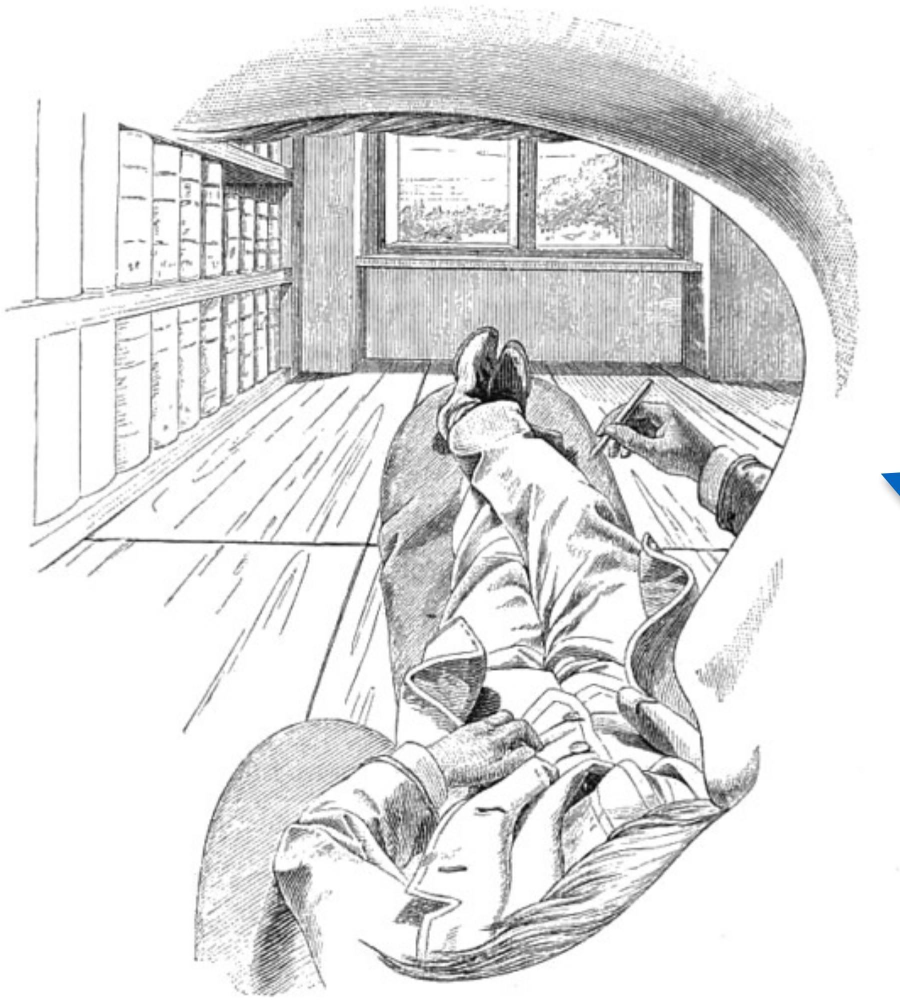
"**self pattern**" (what "I am right now", including observations and memory)

follows from

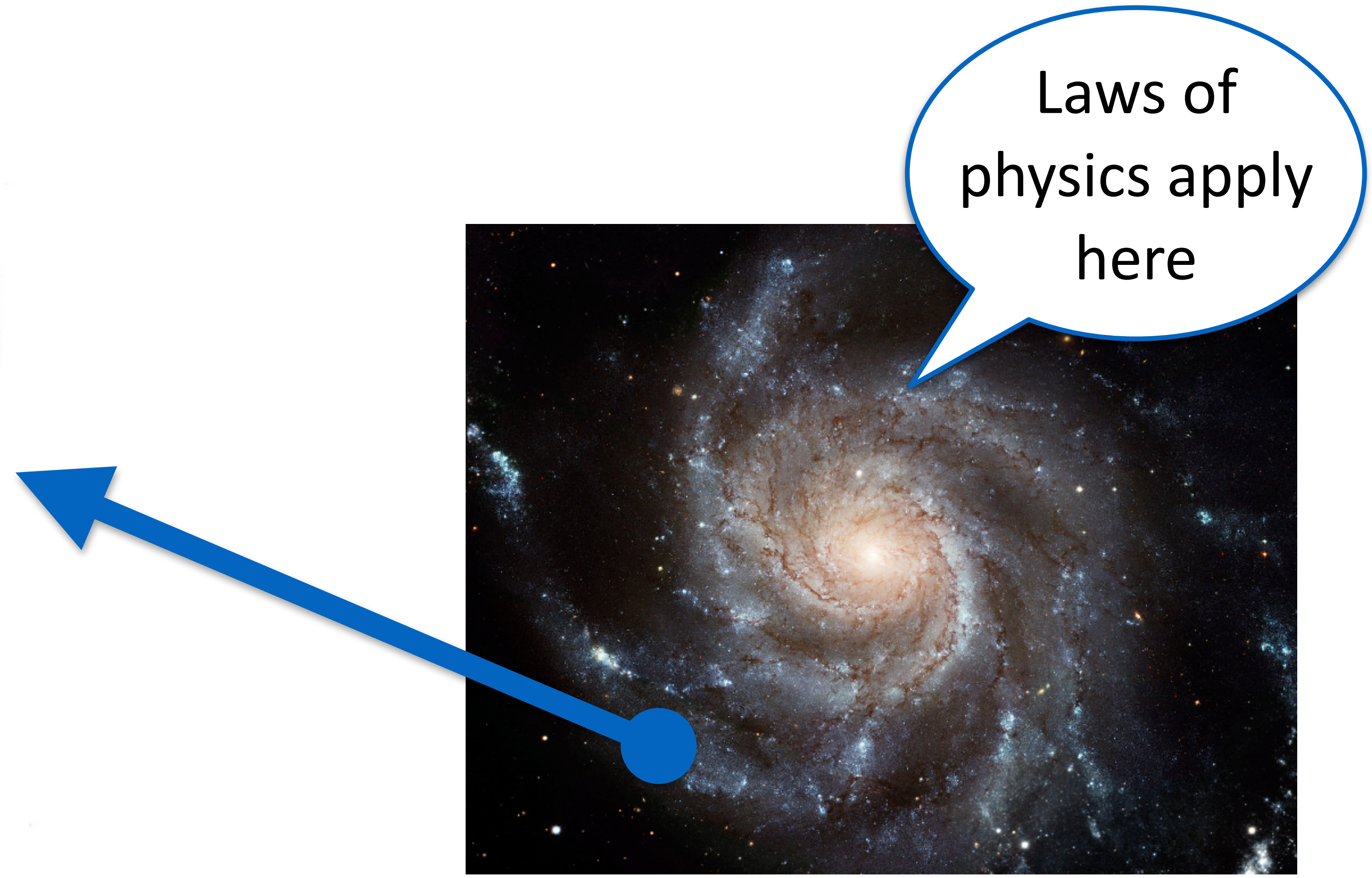
state (and evolution) of the physical world

Problem: methodologically inadequate (recall the exotic puzzles) and conceptually hard to reconcile with quantum theory.

Reversing the standard view



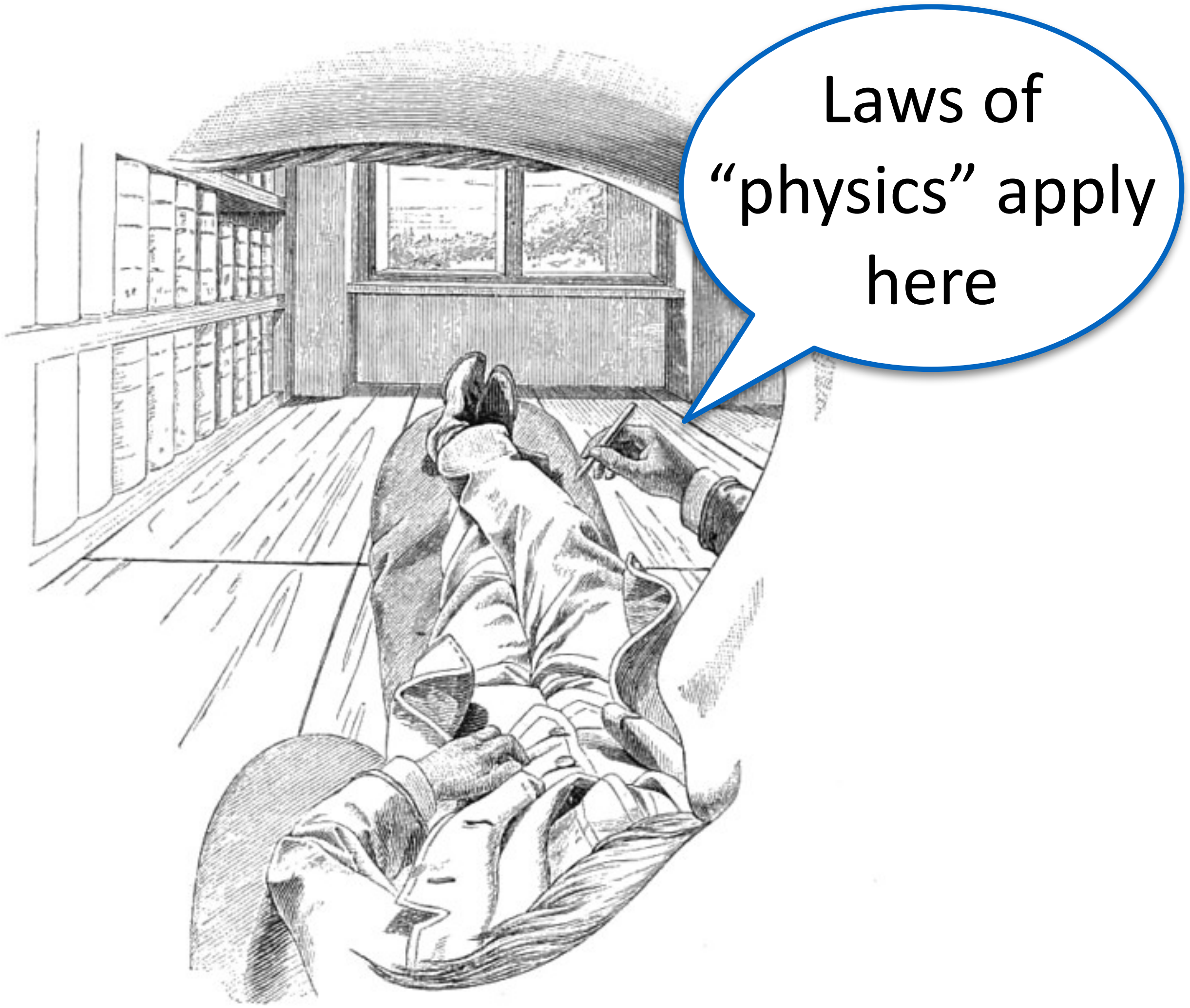
Figur 1.



Reversing the standard view

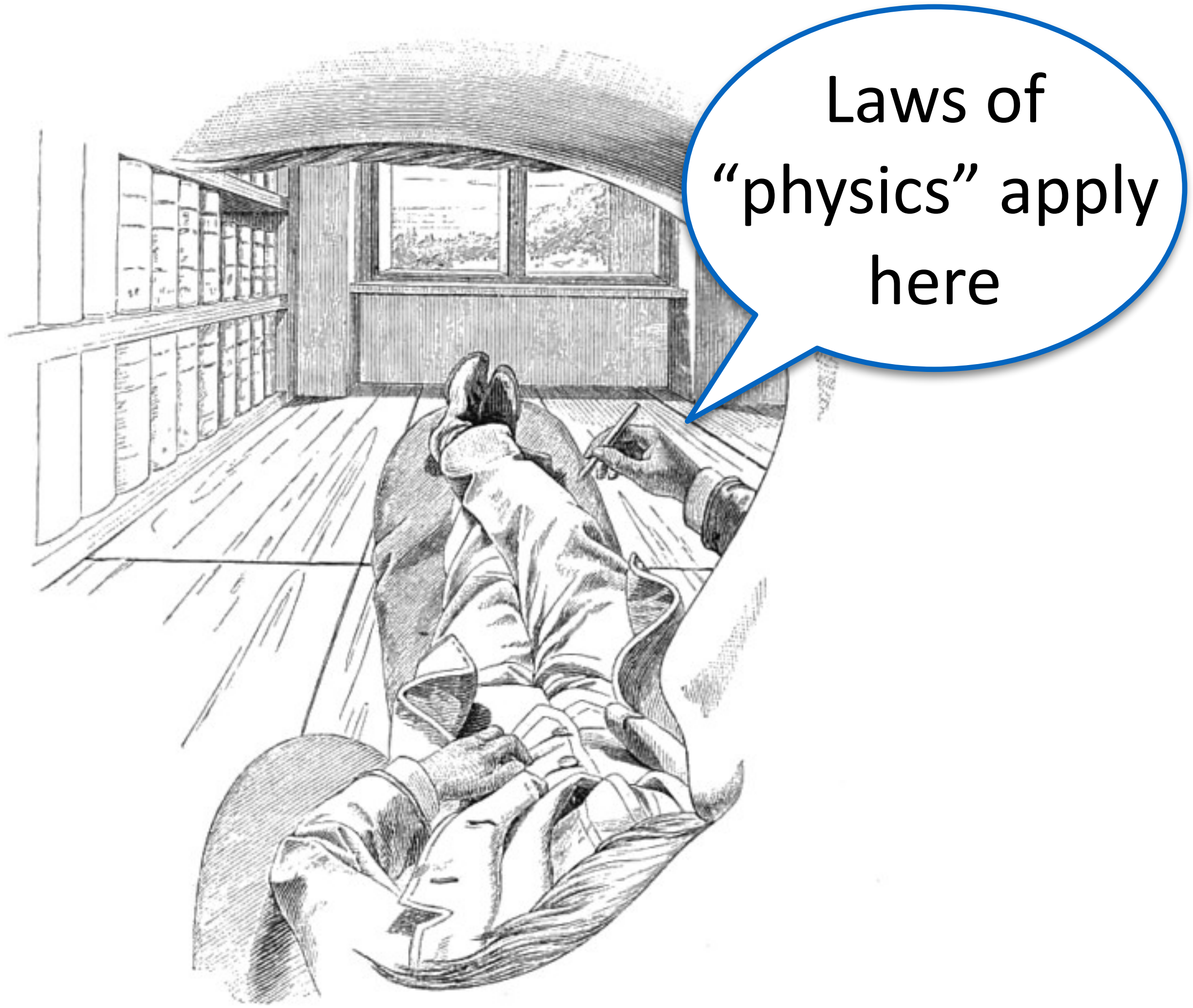
Figur 1.

Reversing the standard view



Figur 1.

Reversing the standard view



Figur 1.

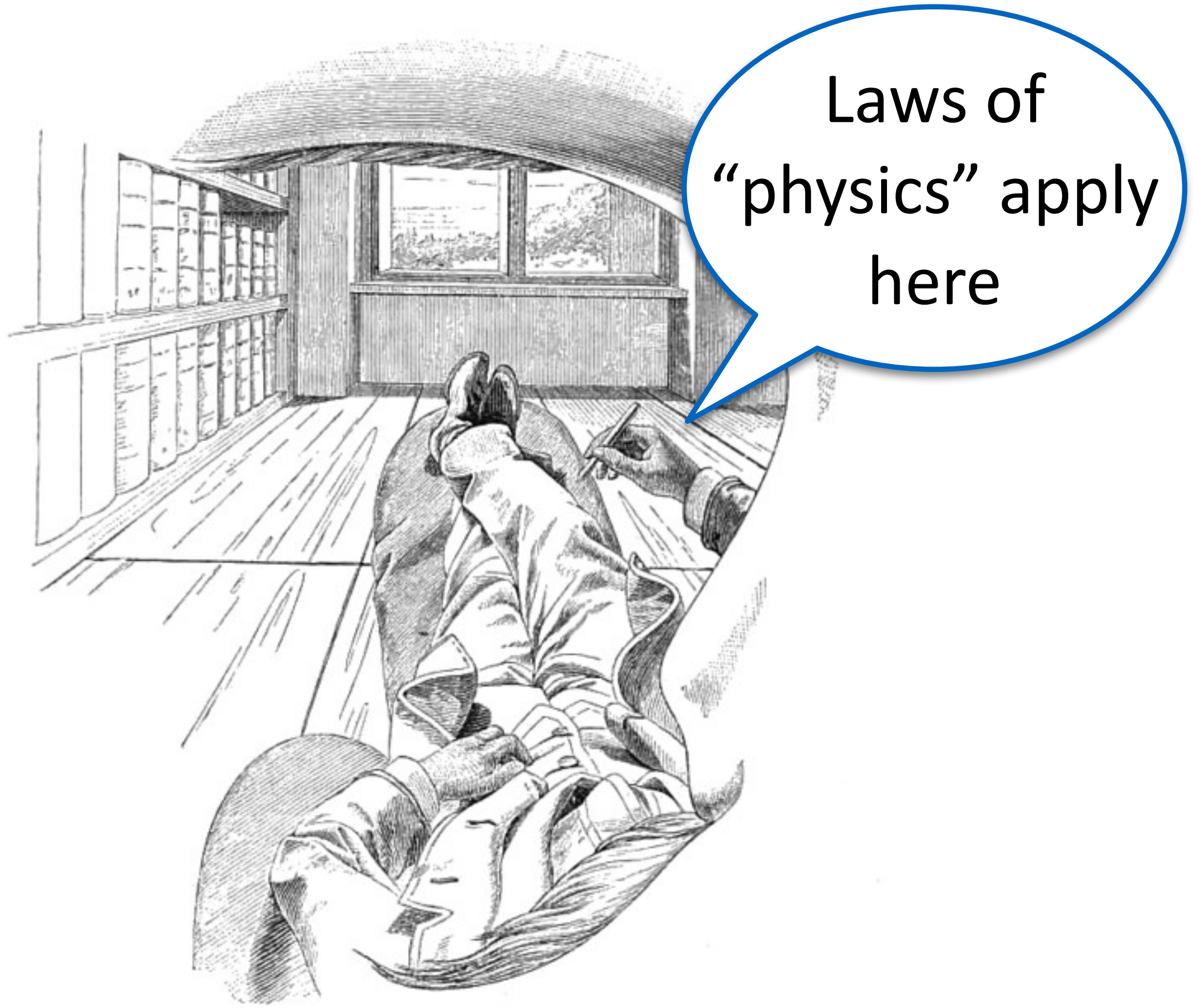
$$\mathbf{P}(y|x)$$

x : self pattern now

y : self pattern next

Universal probability

Reversing the standard view



Figur 1.

$$\mathbf{P}(y|x)$$

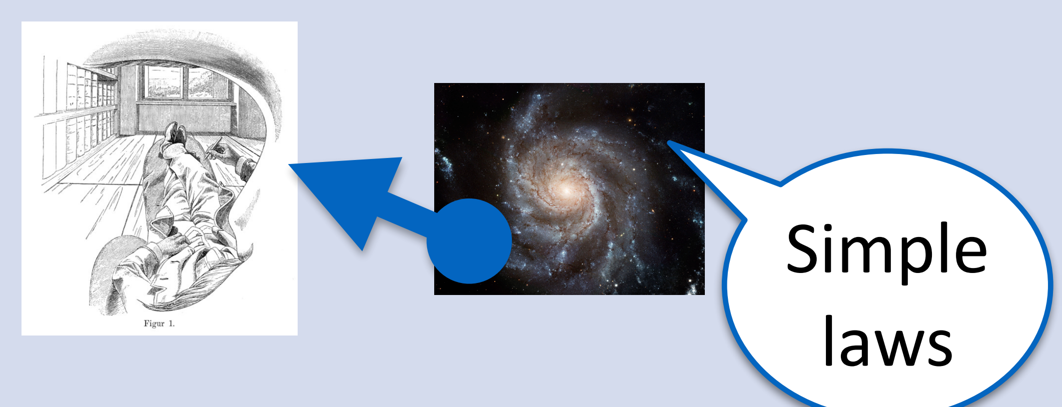
x : self pattern now

y : self pattern next

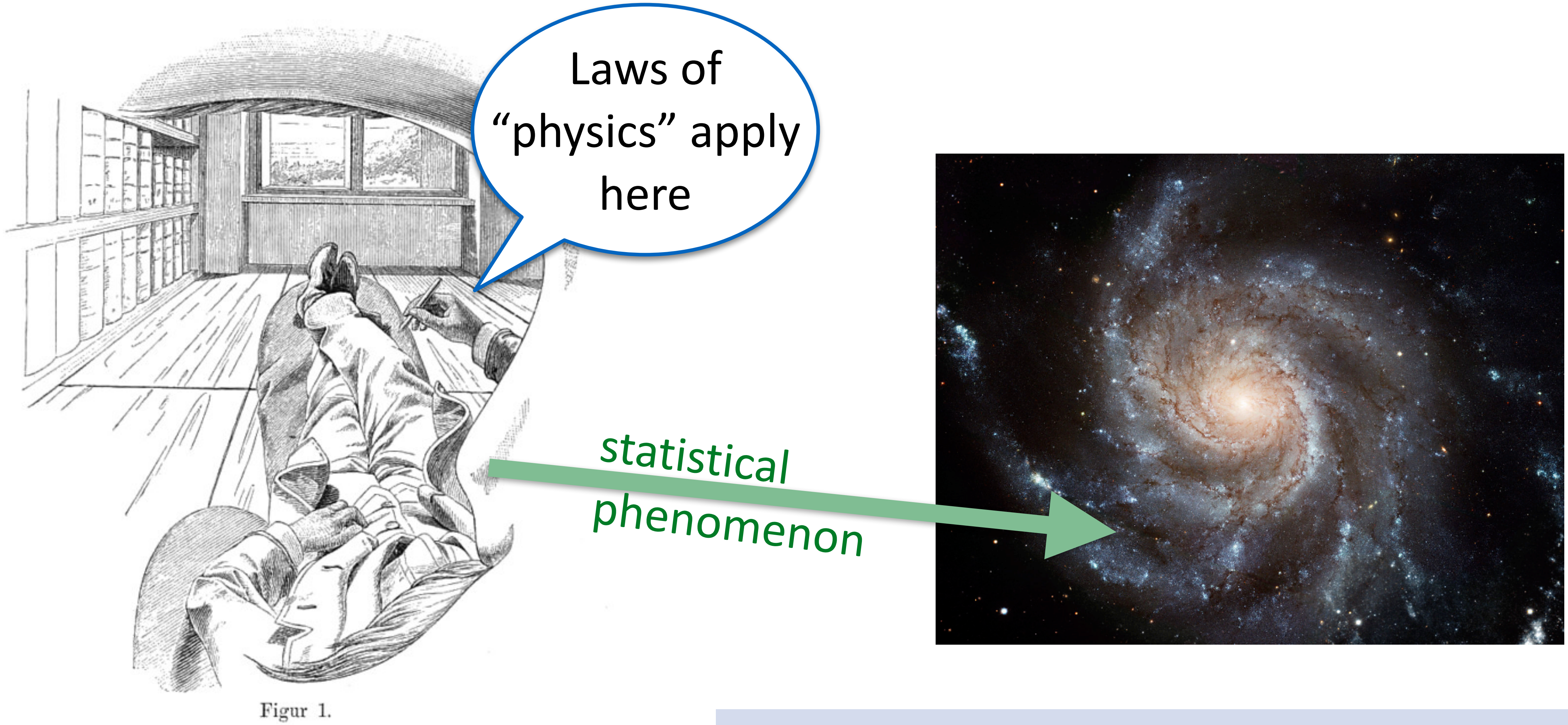
Universal probability

Standard view as an approximation:

In the long run, this will appear as if



Reversing the standard view



$$\mathbf{P}(y|x)$$

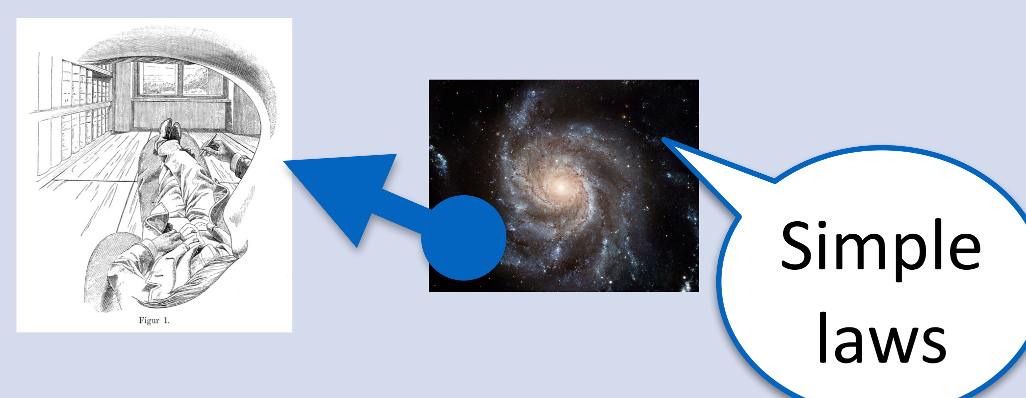
x : self pattern now

y : self pattern next

Universal probability

Standard view as an approximation:

In the long run, this will appear as if



Universal probability

Describe **self patterns** as a finite bit strings, growing one bit at a time.

$$x = 0100101000100000011111.$$

Universal probability

Describe **self patterns** as a finite bit strings, growing one bit at a time.

$$x = 0100101000100000011111.$$

Q: But I’m not a bunch of bits! So **how** do I **encode** my pattern into bits?

A: It doesn’t matter — the resulting theory is **independent of the choice of encoding**, similarly as GR is independent of choice of coordinates.

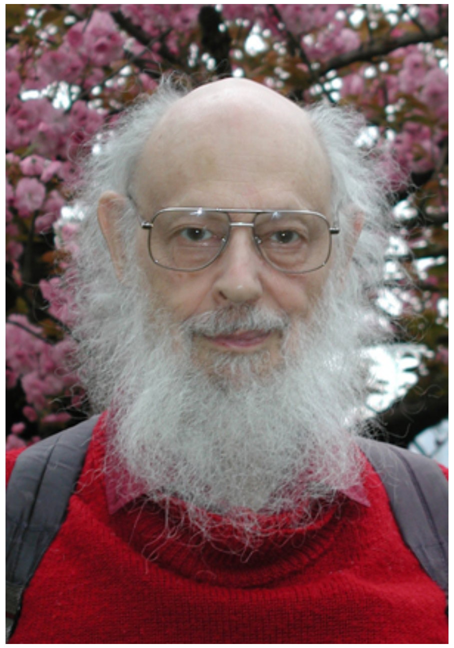
Universal probability

Describe **self patterns** as a finite bit strings, growing one bit at a time.

x = 01001010001000001111.

Q: But I'm not a bunch of bits! So **how** do I **encode** my pattern into bits?

A: It doesn't matter — the resulting theory is **independent of the choice of encoding**, similarly as GR is independent of choice of coordinates.



Ray J. Solomonoff (1964)

Measure: $$\mu(b|x) = \frac{\mu(xb)}{\mu(x)}$$

Prob. that next bit is *b* if now in state *x*.

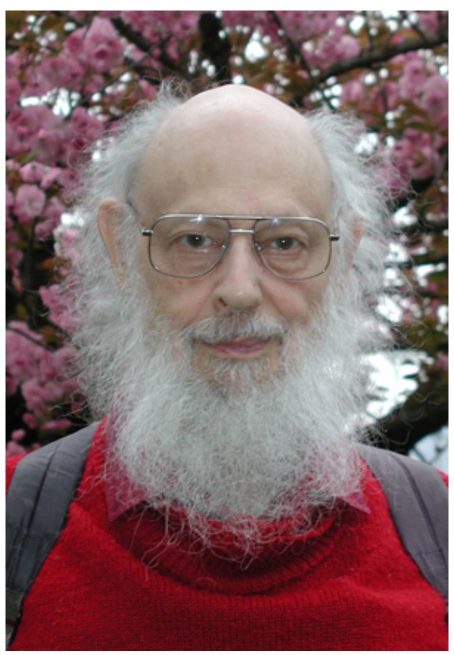
Universal probability

Describe **self patterns** as a finite bit strings, growing one bit at a time.

x = 01001010001000001111.

Q: But I'm not a bunch of bits! So **how** do I **encode** my pattern into bits?

A: It doesn't matter — the resulting theory is **independent of the choice of encoding**, similarly as GR is independent of choice of coordinates.



Ray J. Solomonoff (1964)

Measure: $$\mu(b|x) = \frac{\mu(xb)}{\mu(x)}$$

Prob. that next bit is *b* if now in state *x*.

$$\mu(0|x) + \mu(1|x) = 1.$$

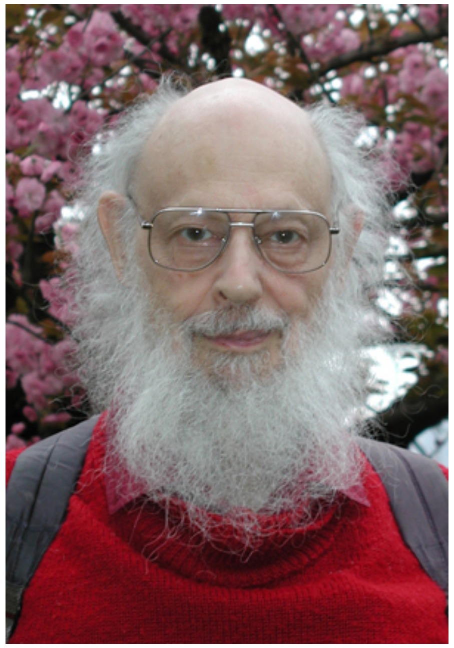
Universal probability

Describe **self patterns** as a finite bit strings, growing one bit at a time.

x = 01001010001000001111.

Q: But I'm not a bunch of bits! So **how** do I **encode** my pattern into bits?

A: It doesn't matter — the resulting theory is **independent of the choice of encoding**, similarly as GR is independent of choice of coordinates.



Ray J. Solomonoff (1964)

Measure: $$\mu(b|x) = \frac{\mu(xb)}{\mu(x)}$$

Prob. that next bit is *b* if now in state *x*.

$$\mu(0|x) + \mu(1|x) = 1.$$

Semimeasure: $$\mu(0|x) + \mu(1|x) \leq 1.$$

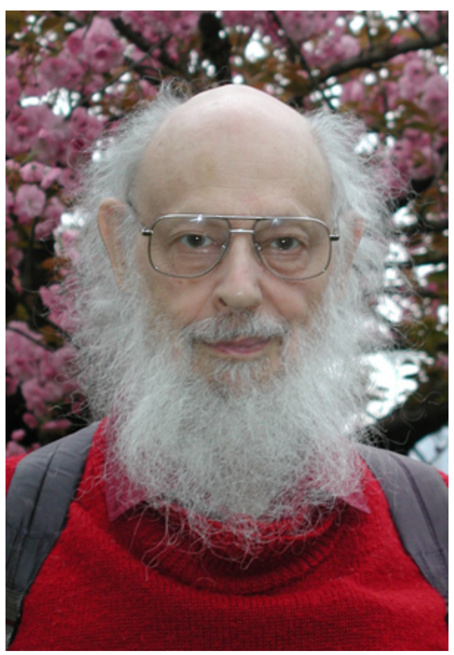
Universal probability

Describe **self patterns** as a finite bit strings, growing one bit at a time.

x = 01001010001000001111.

Q: But I'm not a bunch of bits! So **how** do I **encode** my pattern into bits?

A: It doesn't matter — the resulting theory is **independent of the choice of encoding**, similarly as GR is independent of choice of coordinates.



Measure: $$\mu(b|x) = \frac{\mu(xb)}{\mu(x)}$$

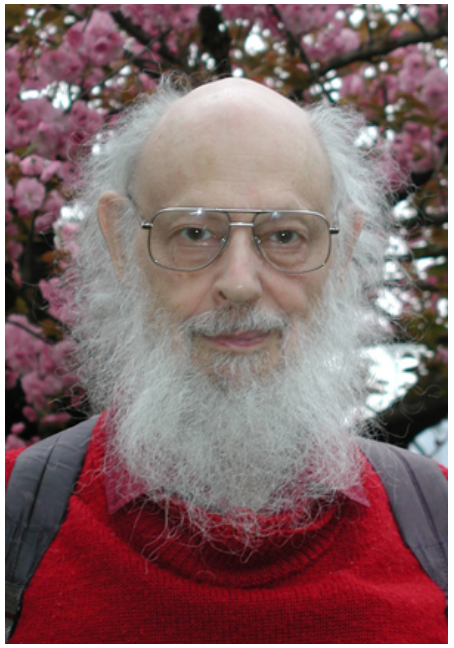
Prob. that next bit is *b* if now in state *x*.

$$\mu(0|x) + \mu(1|x) = 1.$$

Semimeasure: $\mu(0|x) + \mu(1|x) \leq 1.$

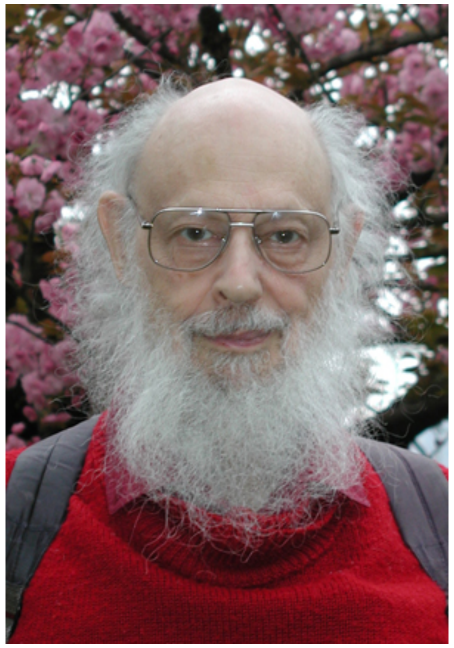
Enumerable semimeasure: there exists an algorithm that, on input *x* and *n*, computes an approx. $\mu_n(x)$ with $\lim_{n\to\infty} \mu_n(x) = \mu(x)$ and $\mu_1 \leq \mu_2 \leq \ldots$

Universal probability



Universal enumerable semimeasure μ**:**

Universal probability

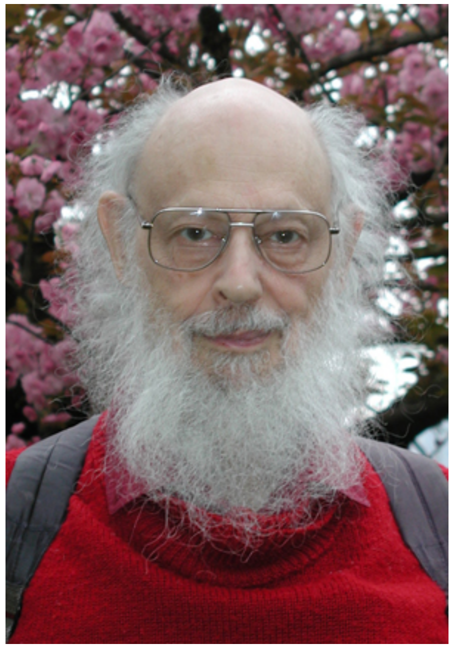


Universal enumerable semimeasure μ:

For every enumerable semimeasure ν there is a constant $c_\nu > 0$ such that $\mu(x) \geq c_\nu \cdot \nu(x)$ for all x.

Basically, a mixture of all enumerable semimeasures.

Universal probability



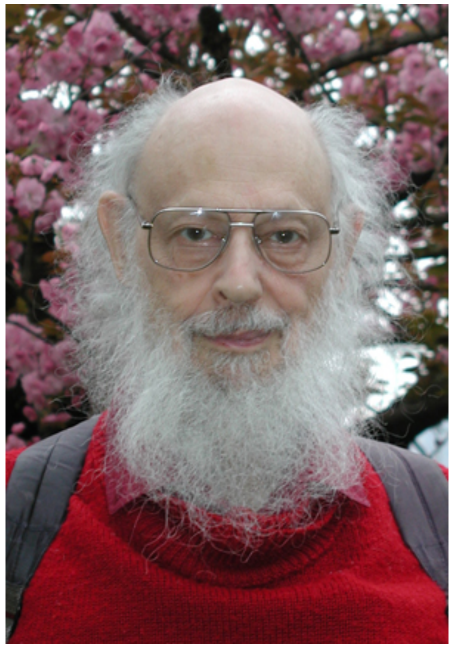
Universal enumerable semimeasure μ:

For every enumerable semimeasure ν there is a constant $c_\nu > 0$ such that $\mu(x) \geq c_\nu \cdot \nu(x)$ for all x.

Basically, a mixture of all enumerable semimeasures.

normalize it ⟶ **universal probability P**.

Universal probability



Universal enumerable semimeasure μ:

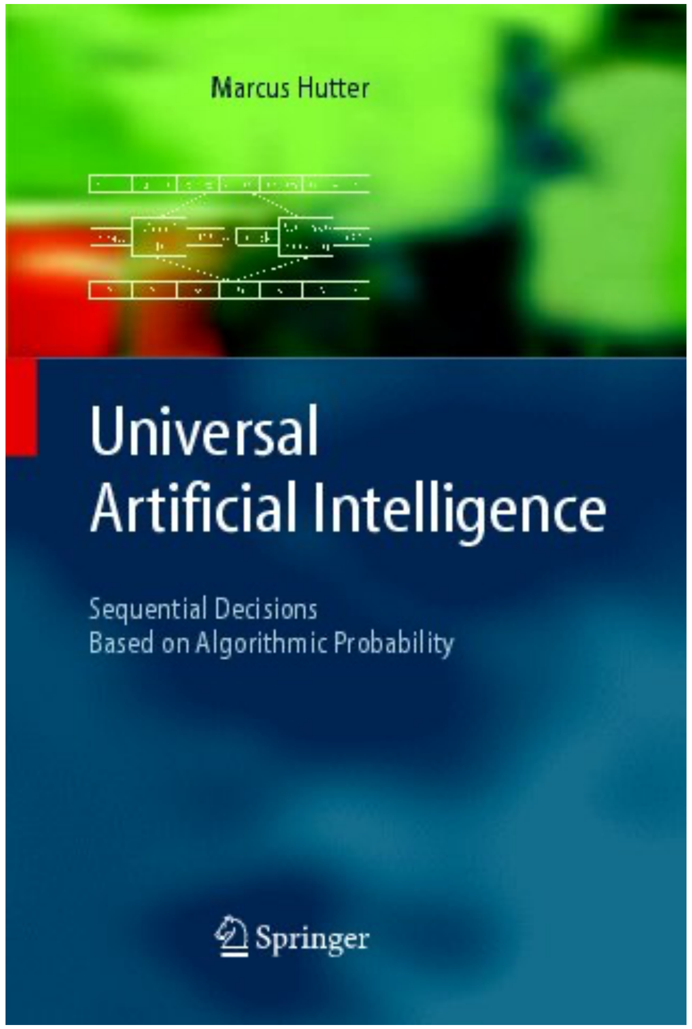
For every enumerable semimeasure ν there is a constant $c_\nu > 0$ such that $\mu(x) \geq c_\nu \cdot \nu(x)$ for all x.

Basically, a mixture of all enumerable semimeasures.

normalize it ⟶ **universal probability P**.

Application elsewhere (not in my approach):

- Gives higher probability to simpler bit strings (i.e. generated by shorter programs). **Occam's razor**.
- Uncomputable, but in principle useful for **induction** ⟶ "Universal Artificial Intelligence"
- **Solomonoff induction**: yields provably correct predictions asymptotically (quickly) in all computable environments.



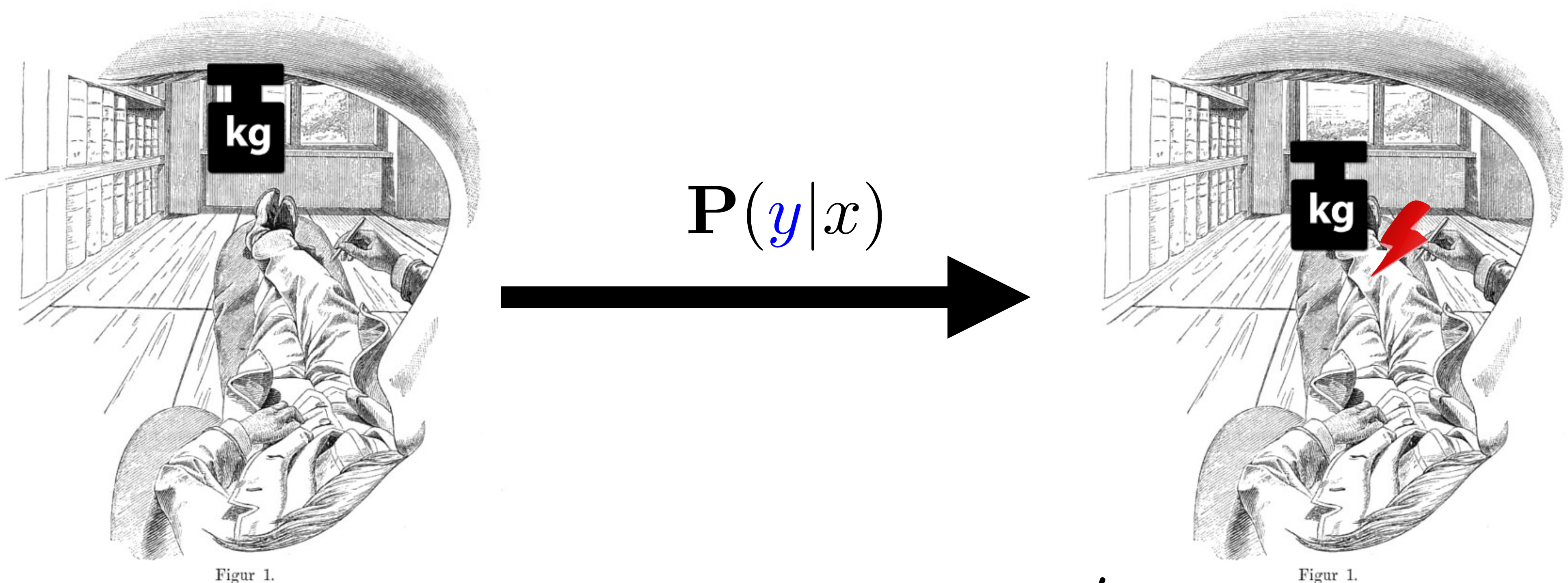
Postulates of an (incomplete) idealist theory

Postulates of an (incomplete) idealist theory

At every (subjective) moment, “I” am a self-pattern *x*, and a couple of moments later, I will be a self-pattern xy, with universal probability **P**(y|x).

Postulates of an (incomplete) idealist theory

At every (subjective) moment, "I" am a self-pattern *x*, and a couple of moments later, I will be a self-pattern xy, with universal probability **P**(y|x).

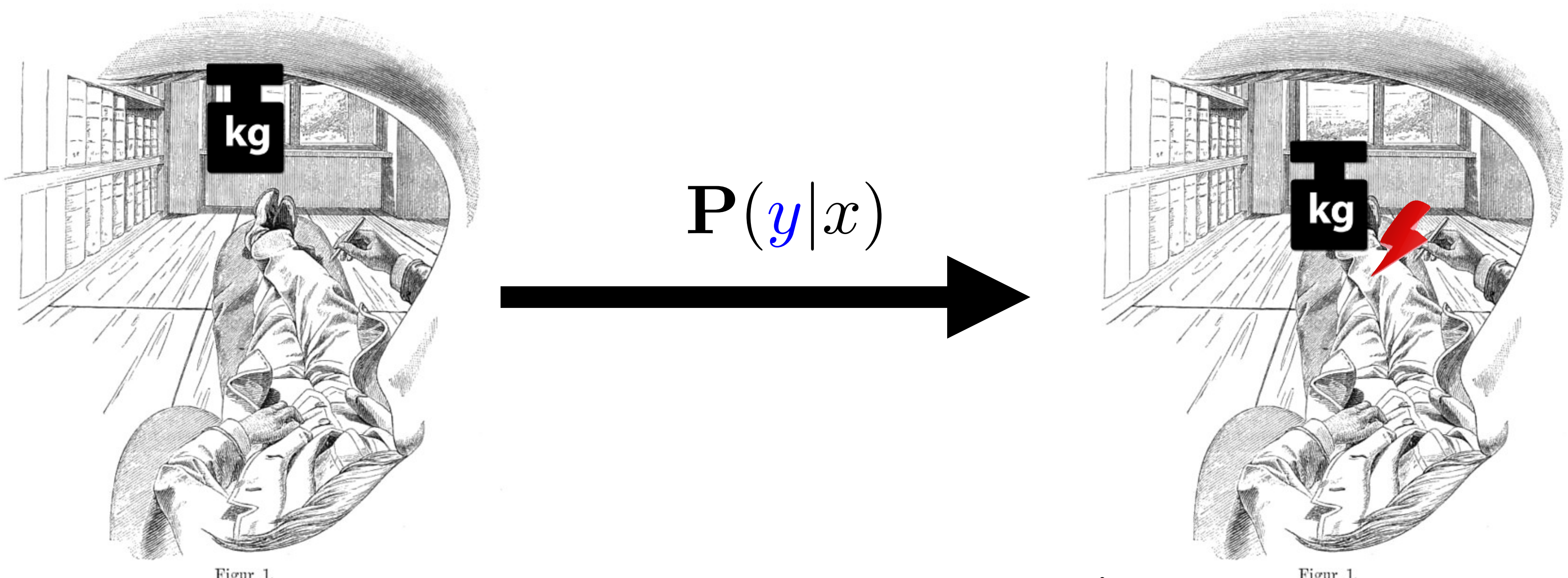


$x = 1001$

$x' = 1001\color{blue}{1111101} = x\color{blue}{y}$

Postulates of an (incomplete) idealist theory

At every (subjective) moment, "I" am a self-pattern *x*, and a couple of moments later, I will be a self-pattern xy, with universal probability **P**(y|x).

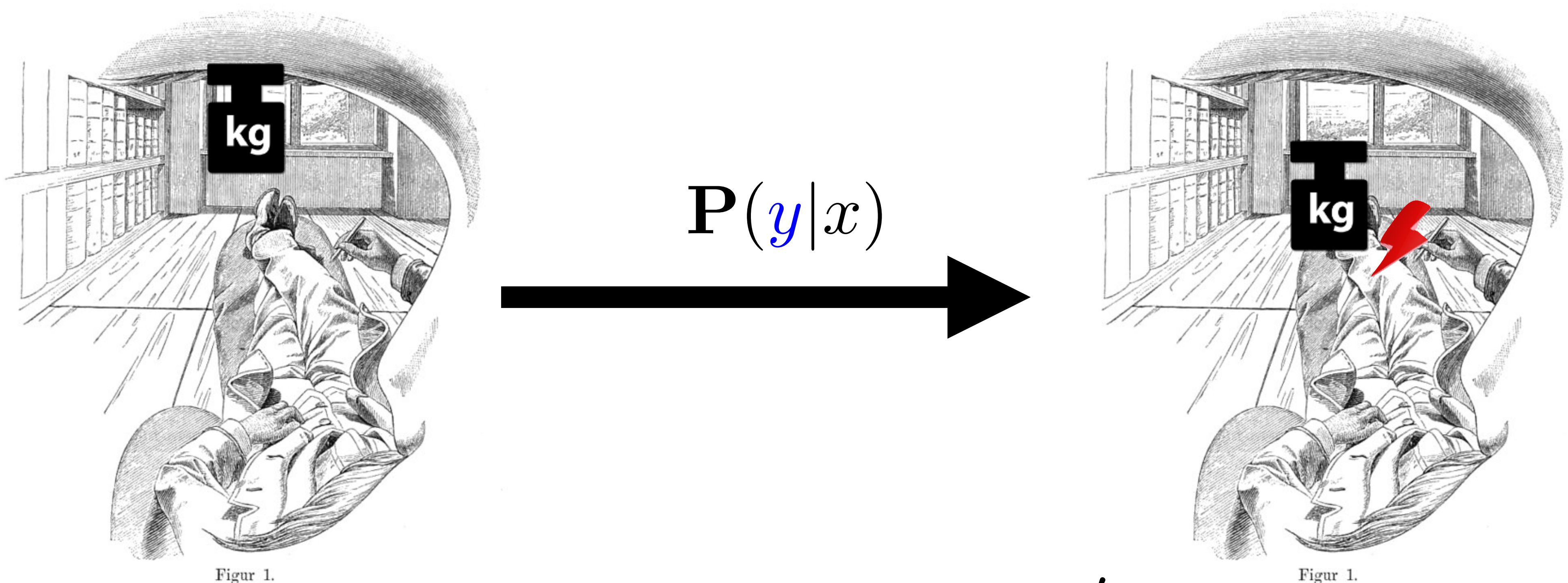


$x = 1001$ $\qquad$ $x' = 1001\textcolor{blue}{1111101} = x\textcolor{blue}{y}$

This is a fundamental, objective, private chance that does not arise from any lack of knowledge, or any "external world" in which my pattern would be embedded. **"I am an unembedded pattern".**

Postulates of an (incomplete) idealist theory

At every (subjective) moment, "I" am a self-pattern *x*, and a couple of moments later, I will be a self-pattern xy, with universal probability **P**(y|x).

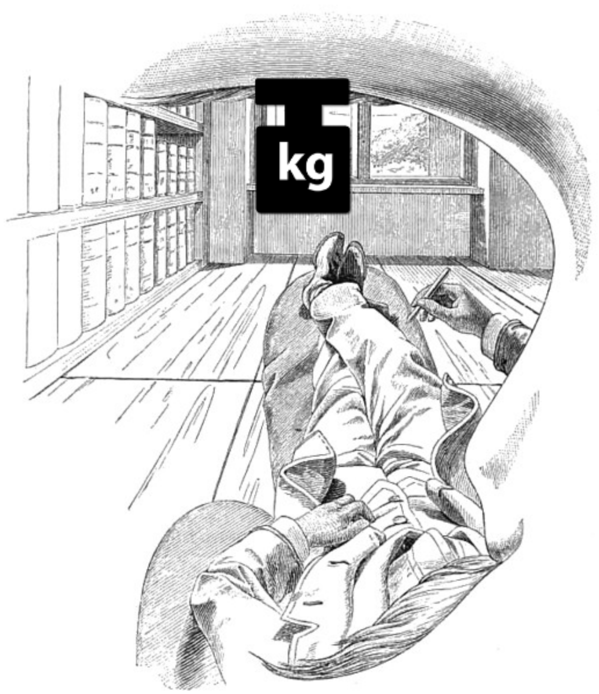


$x = 1001$ $\qquad$ $x' = 1001\color{blue}{1111101} = x\color{blue}{y}$

This is a fundamental, objective, private chance that does not arise from any lack of knowledge, or any "external world" in which my pattern would be embedded. **"I am an unembedded pattern".**

(Incomplete theory, because "forgetting" not yet treated.)

Consistency with standard physics

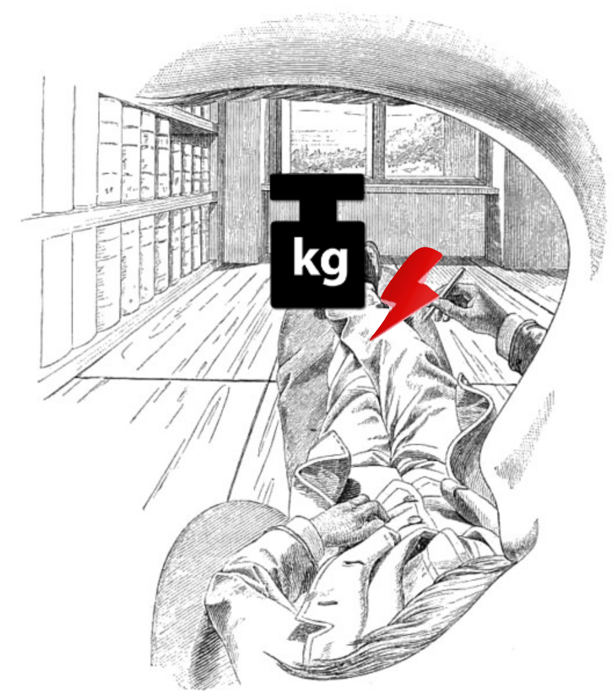


Figur 1.

$x = 1001$

$\mathbf{P}(y|x)$

universal probability

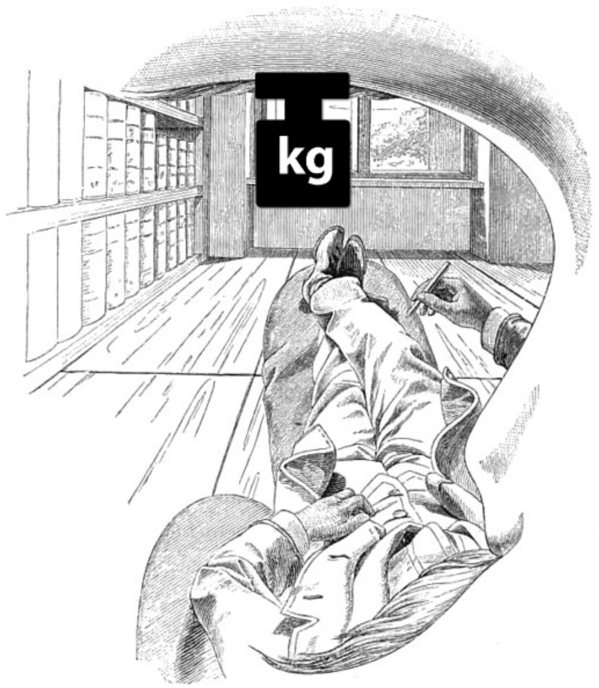


Figur 1.

$x' = 1001\color{blue}{1111101} = x\color{blue}{y}$

Consistency with standard physics

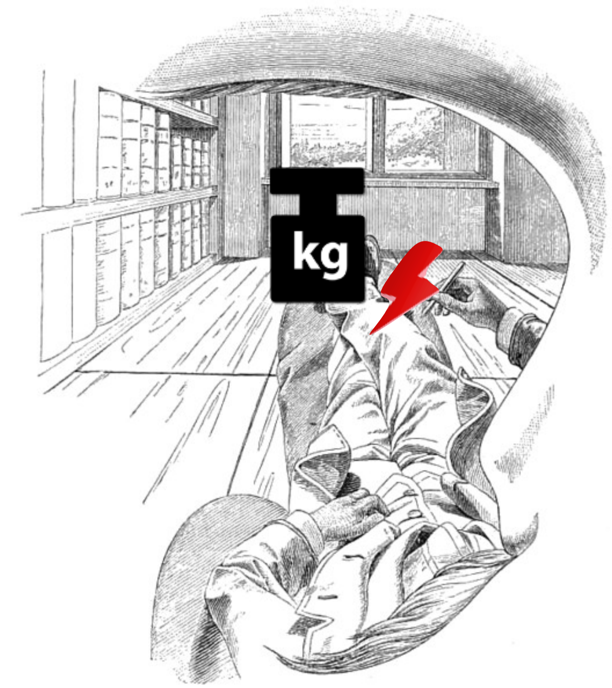
The standard view would tell us to compute the "physical" probability $\mathbf{P}_{\text{phys}}(y|x)$ arising from, say, the wave function of the universe.



$x = 1001$

$\mathbf{P}(y|x)$

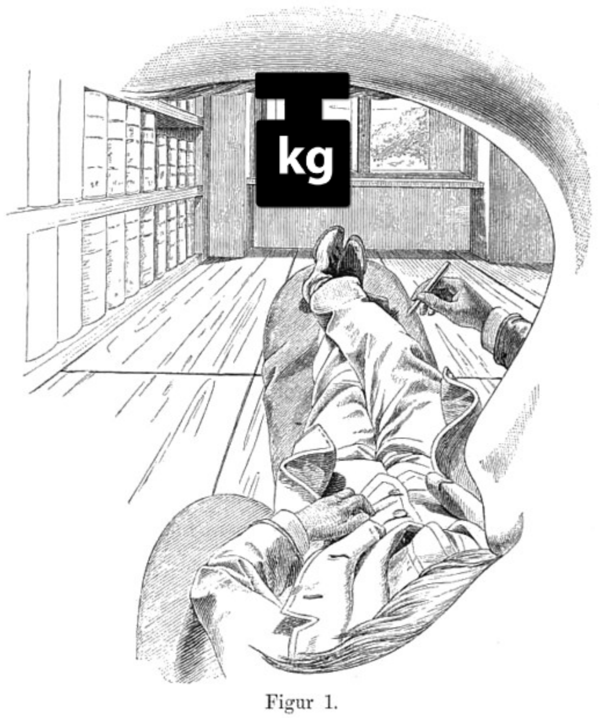
universal probability



$x' = 1001\,1111101 = xy$

Consistency with standard physics

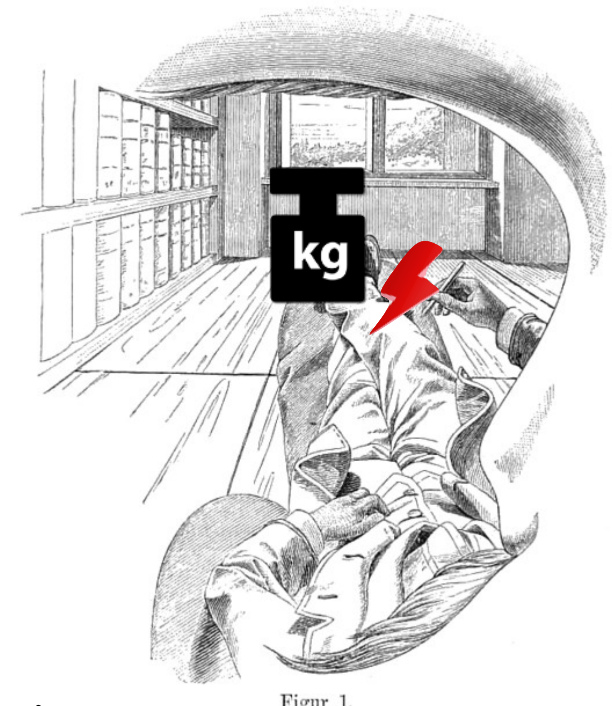
The standard view would tell us to compute the "physical" probability $\mathbf{P}_{\mathrm{phys}}(y|x)$ arising from, say, the wave function of the universe.



$x = 1001$

$\mathbf{P}(y|x)$

universal probability



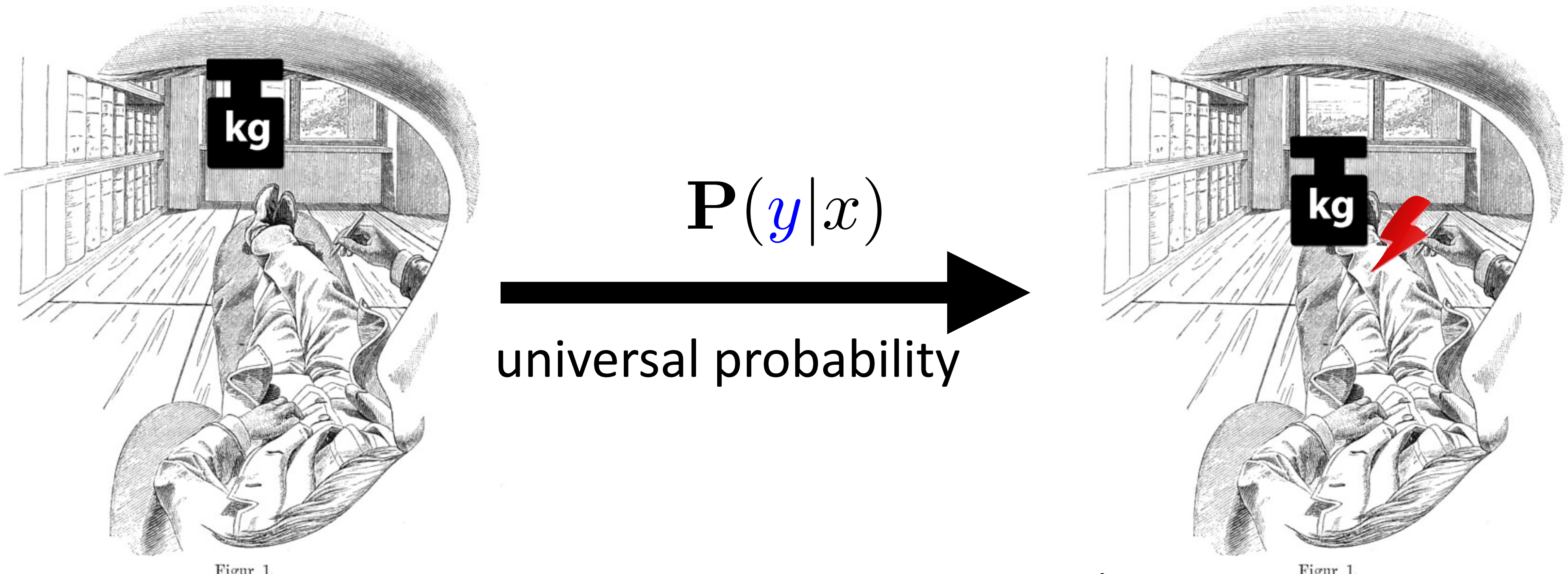
$x' = 1001\textcolor{blue}{1111101} = xy$

Theorem. In the limit of a large number $n = \ell(x)$ of self-pattern bits,

$$|\mathbf{P}(y|x) - \mathbf{P}_{\mathrm{phys}}(y|x)| \overset{n\to\infty}{\longrightarrow} 0.$$

Consistency with standard physics

The standard view would tell us to compute the "physical" probability $\mathbf{P}_{\mathrm{phys}}(y|x)$ arising from, say, the wave function of the universe.



$x = 1001$ $\qquad$ $x' = 1001111111101 = xy$

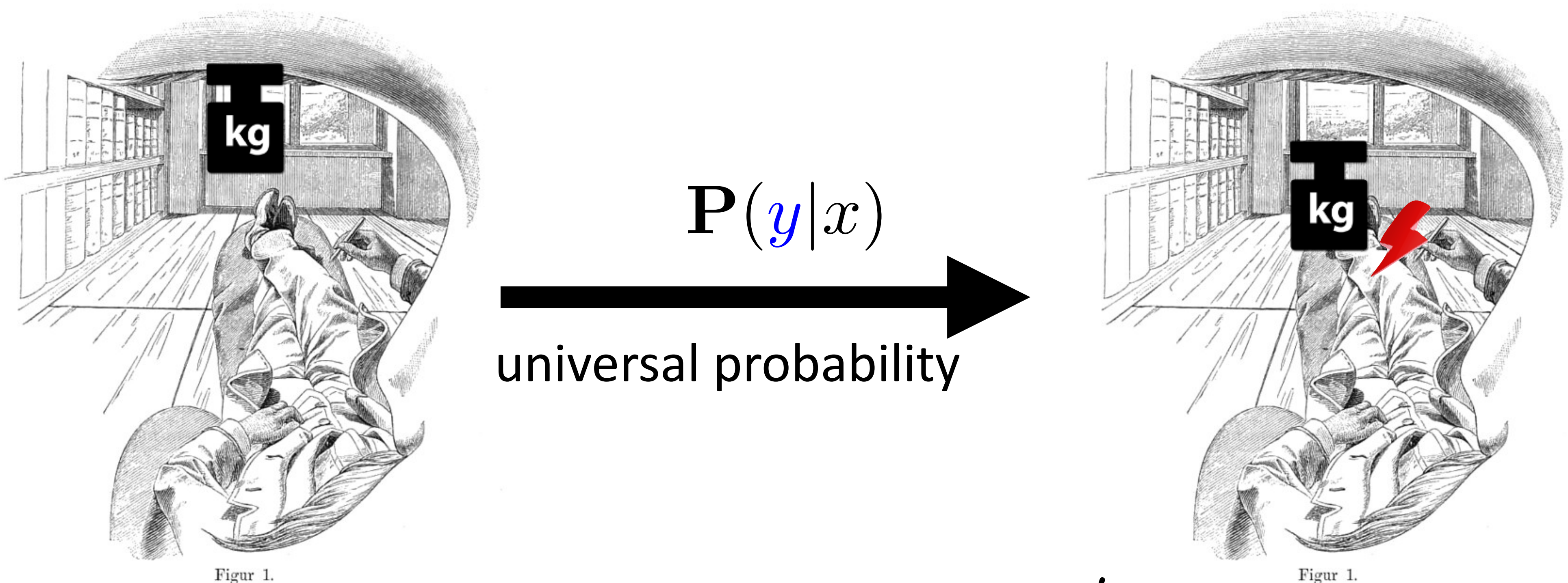
Theorem. In the limit of a large number $n = \ell(x)$ of self-pattern bits,

$$|\mathbf{P}(y|x) - \mathbf{P}_{\mathrm{phys}}(y|x)| \stackrel{n\to\infty}{\longrightarrow} 0.$$

Proof. Physical versions of the Church-Turing thesis
$\Rightarrow \mathbf{P}_{\mathrm{phys}}$ is in principle computable. Thus, due to Solomonoff's universal induction, convergence above happens with $\mathbf{P}_{\mathrm{phys}}$-prob. 1.

Consistency with standard physics

The standard view would tell us to compute the "physical" probability $\mathbf{P}_{\mathrm{phys}}(y|x)$ arising from, say, the wave function of the universe.



$x = 1001$ $\qquad\qquad$ $x' = 1001\color{blue}{111101} = x\color{blue}{y}$

Interpretation. If the self-pattern contains enough information on the (for me) relevant aspects of the physical world, then universal probability will "detect" these regularities (**Solomonoff induction**) and assign high probability to the fact that these regularities will remain present.

Hence, physical and universal probabilities will agree in their predictions.

Consistency with standard physics

Consistency with standard physics is good news, but it is **not enough**.

Consistency with standard physics

Consistency with standard physics is good news, but it is **not enough**.

Now that I hold a large amount of information on a (possible) external physical world, universal probability predicts chances that conform with that (possible) external world in the future. Fair enough.
But why should I get there in the first place if universal probability is all there is, and no external world is assumed to begin with?

Consistency with standard physics

Consistency with standard physics is good news, but it is **not enough**.

Now that I hold a large amount of information on a (possible) external physical world, universal probability predicts chances that conform with that (possible) external world in the future. Fair enough.
But why should I get there in the first place if universal probability is all there is, and no external world is assumed to begin with?

As we will now show, universal probability predicts an "external world".

Consistency with standard physics

Consistency with standard physics is good news, but it is **not enough**.

Now that I hold a large amount of information on a (possible) external physical world, universal probability predicts chances that conform with that (possible) external world in the future. Fair enough.
But why should I get there in the first place if universal probability is all there is, and no external world is assumed to begin with?

As we will now show, universal probability predicts an "external world".

This does **not** make it a "theory of everything" because it cannot predict most properties of that world.

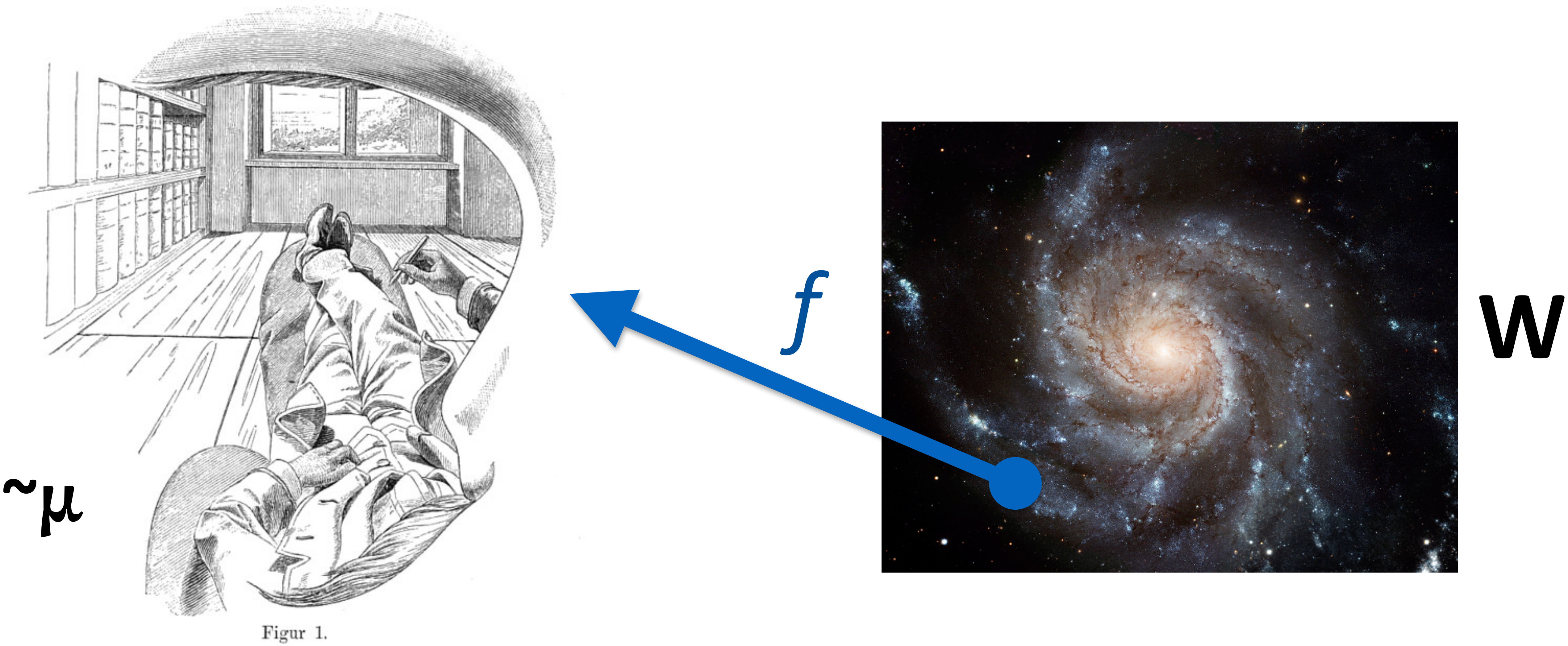
Candidate external worlds

Candidate external worlds

Def.: A **computational ontological model for μ** is a stochastic process ("world" W) that can in principle be run on a probabilistic Turing machine, together with a computable bit-string-valued random variable *f ("locates / reads the self-pattern from world W")* yielding self-patterns evolving as described by **μ**.

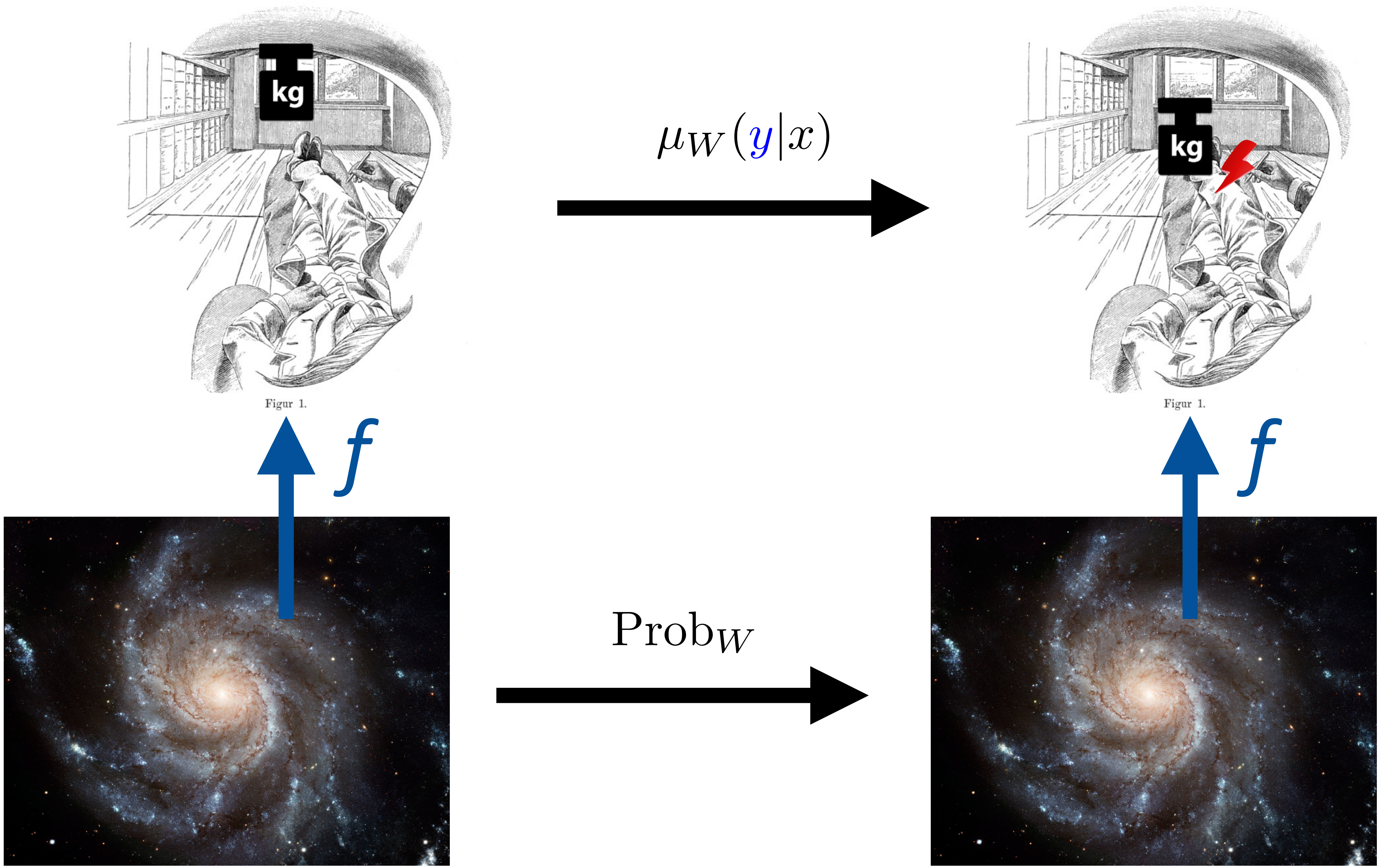
Candidate external worlds

Def.: A **computational ontological model for μ** is a stochastic process ("world" W) that can in principle be run on a probabilistic Turing machine, together with a computable bit-string-valued random variable *f ("locates / reads the self-pattern from world W")* yielding self-patterns evolving as described by **μ**.



Computational ontological models

Basically, this formalizes the “standard view”.

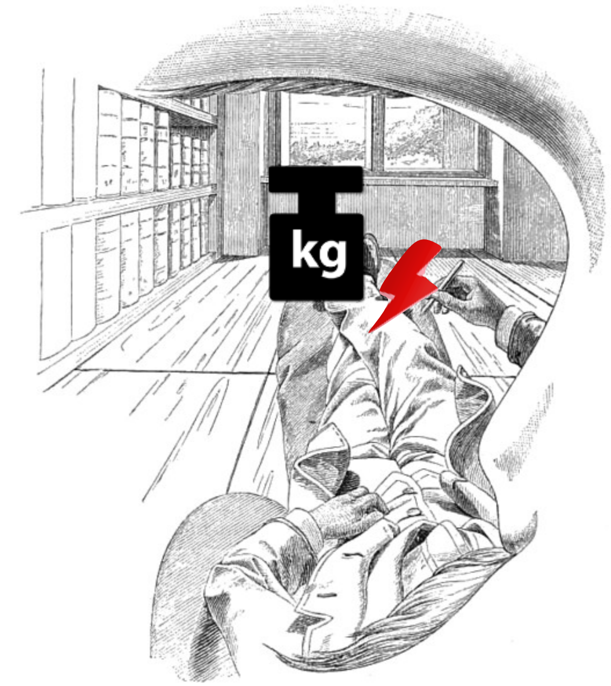


An emergent notion of **external world**

Figur 1.

$$\mathbf{P}(y|x)$$

$$\approx \mu_W(y|x)$$



Figur 1.

Theorem: *Before* agent holds any information (or *after loosing* all info), there is universal probability **P** of at least

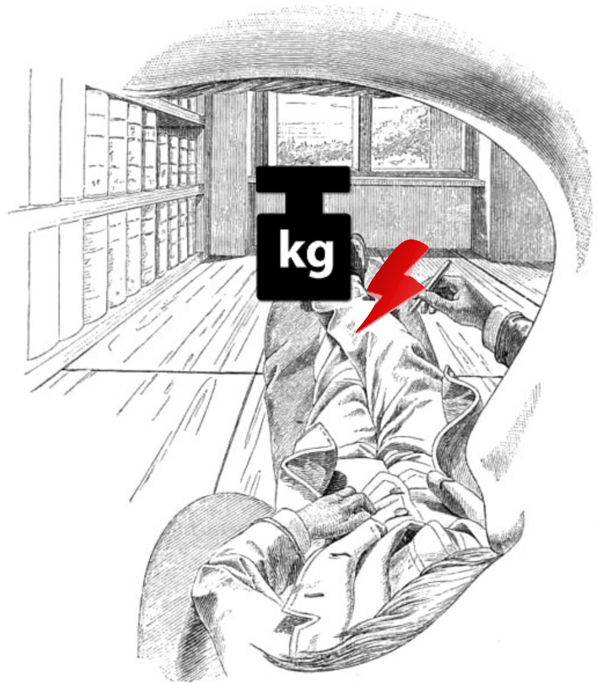
$$2^{-\mathrm{K}(W)}$$

that ontological model ("world") W is seen in the long run, i.e. that $|\mathbf{P}(y|x_1,\ldots,x_n) - \mu_W(y|x_1,\ldots,x_n)| \longrightarrow 0.$

An emergent notion of **external world**

Figur 1.

$$\mathbf{P}(y|x) \longrightarrow \approx \mu_W(y|x)$$



Figur 1.

Theorem: *Before* agent holds any information (or *after loosing* all info), there is universal probability **P** of at least

$$2^{-\mathrm{K}(W)}$$

← description length of *W* on a universal computer

that ontological model ("world") W is seen in the long run, i.e. that $|\mathbf{P}(y|x_1,\ldots,x_n) - \mu_W(y|x_1,\ldots,x_n)| \longrightarrow 0.$

actual chances according to univ. probability

chances as determined by world *W*.

An emergent notion of **external world**

Properties of this (**probabilistic**) world W:

- K(*W*) probably small: W has **simple** “laws of nature”.
- **Actual realization** seen by agent typically **complex** (compare: coin toss).
- In particular, μ_W is probabilistically **computable** (recall: **P** isn’t!)
- Such processes typically start in a **state of low entropy**. Big bang?

Broadly consistent with what we observe!

Theorem: *Before* agent holds any information (or *after loosing* all info),
there is universal probability **P** of at least

$$2^{-\mathrm{K}(W)}$$

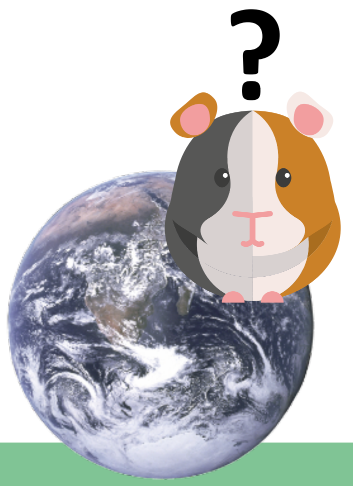
description length of *W* on a universal computer

that ontological model (“world”) W is seen in the long run,
i.e. that $|\mathbf{P}(y|x_1,\ldots,x_n) - \mu_W(y|x_1,\ldots,x_n)| \longrightarrow 0.$

actual chances according to univ. probability

chances as determined by world *W*.

Outline



1. **Conceptual puzzles**
   … that challenge the standard view.

2. **Sketch of an idealist (toy) theory**
   … "self" fundamental, external world emergent.

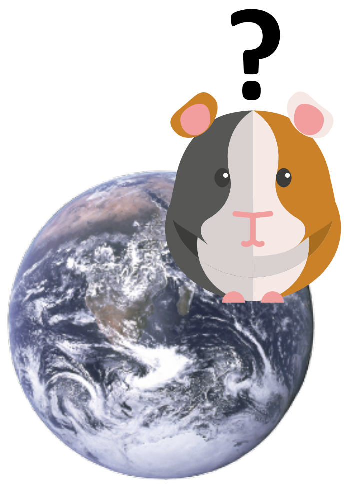
3. **Objective reality as a emergent approximation**
   … probabilistic zombies, and other surprises.

4. **Example: dissolution of the Boltzmann brain problem**

Outline

1. **Conceptual puzzles**

   … that challenge the standard view.



2. **Sketch of an idealist (toy) theory**

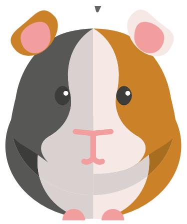
   … "self" fundamental, external world emergent.

3. **Objective reality as a emergent approximation**

   … probabilistic zombies, and other surprises.

4. **Example: dissolution of the Boltzmann brain problem**

An emergent notion of **objective reality**

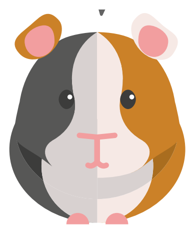


Alice

the guinea pig

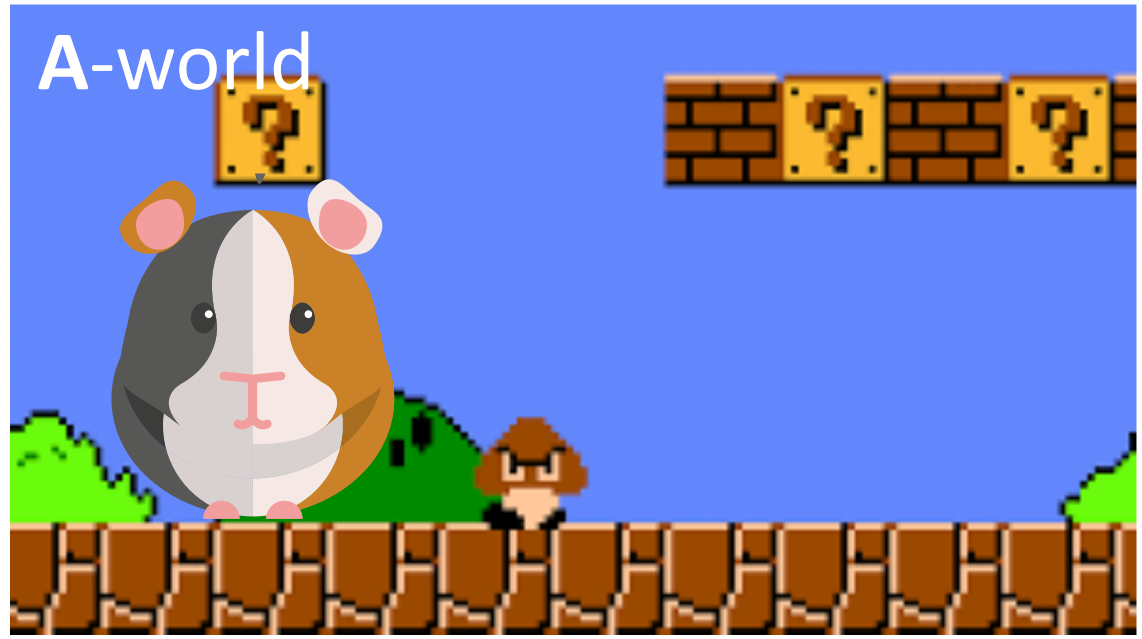
… goes through patterns that look like…

An emergent notion of **objective reality**

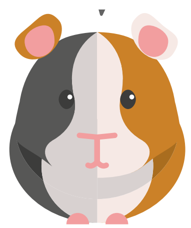


Alice
the guinea pig

… goes through patterns that look like…



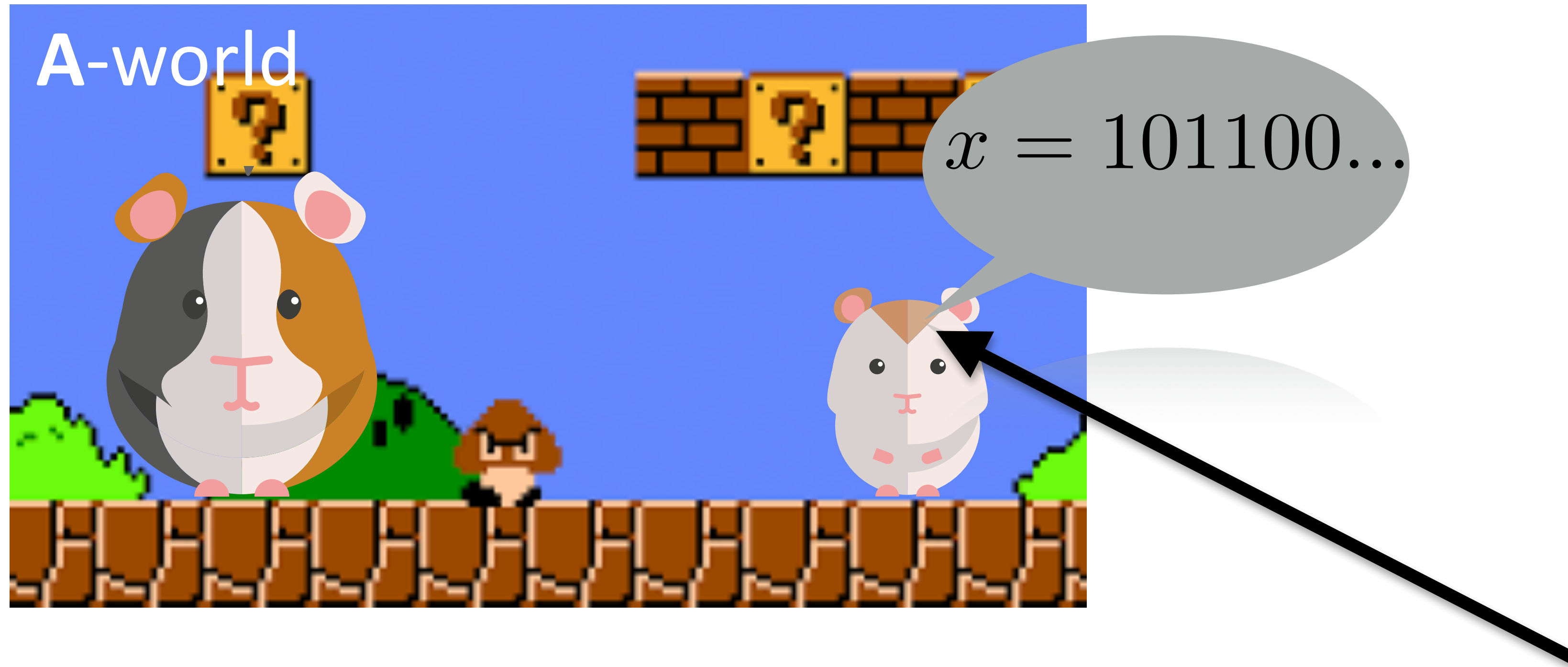
An emergent notion of **objective reality**



Alice

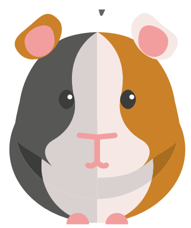
the guinea pig

… goes through patterns that look like…



Suppose in **A**-world, there is another bit-string valued random variable, **B**.

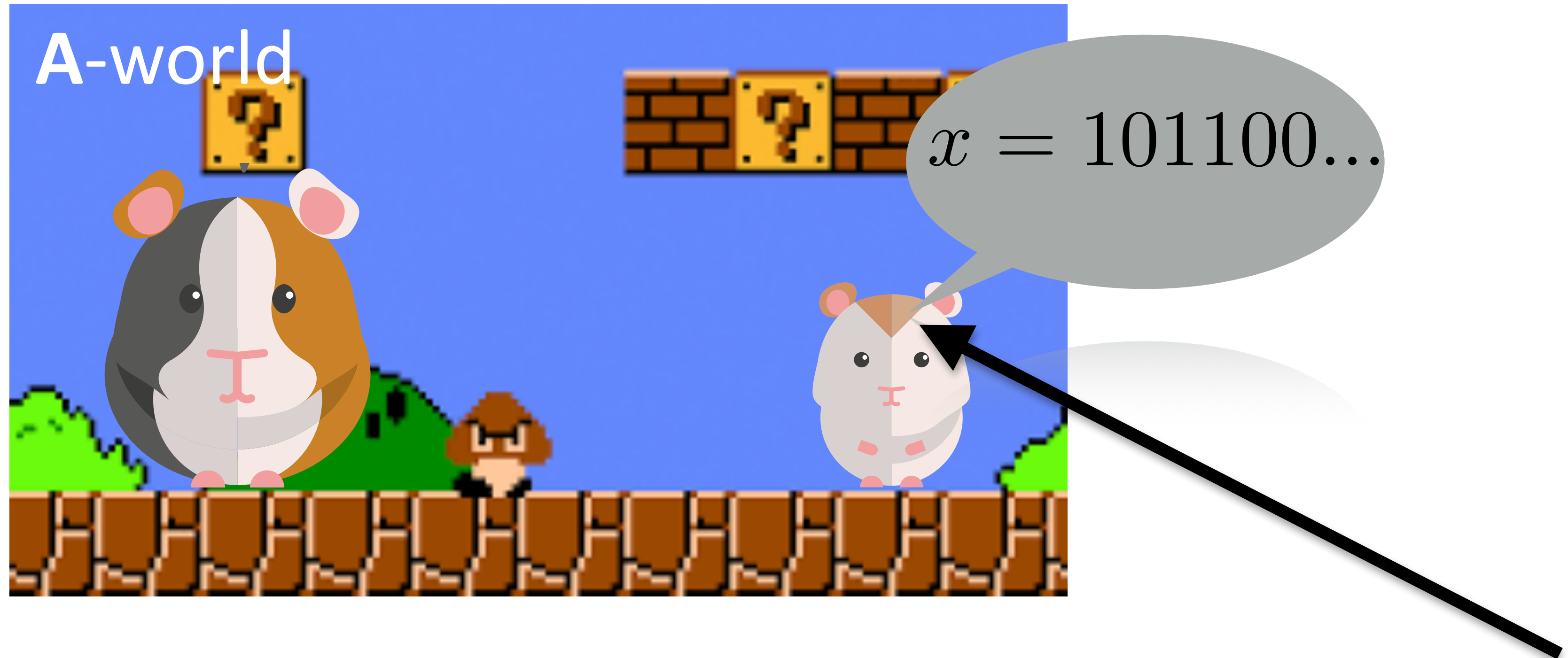
An emergent notion of **objective reality**



Alice

the guinea pig

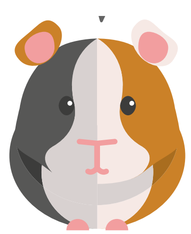
… goes through patterns that look like…



Suppose in **A**-world, there is another bit-string valued random variable, **B**.

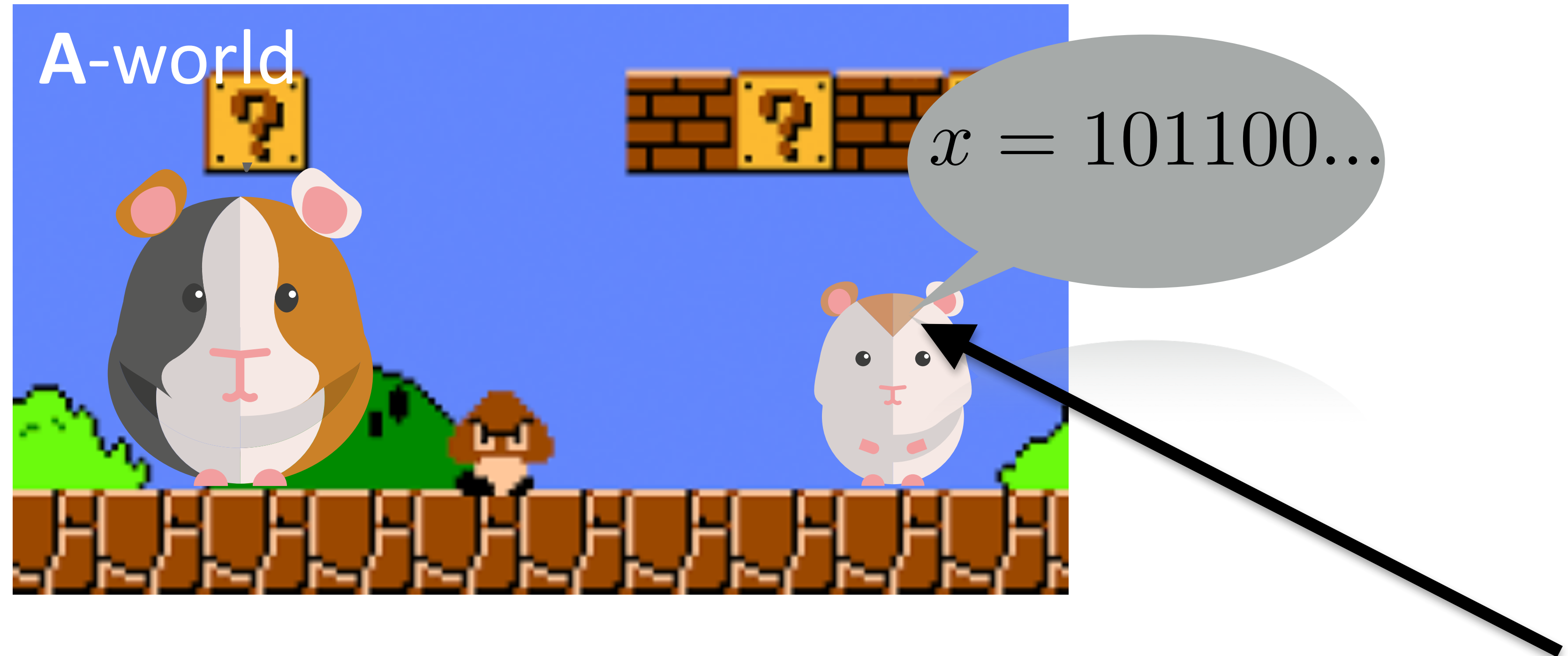
Does **B** faithfully represent some first-person perspective?

An emergent notion of **objective reality**



Alice
the guinea pig

… goes through patterns that look like…



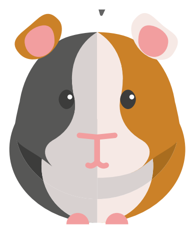
Suppose in **A**-world, there is another bit-string valued random variable, **B**.

Does **B** faithfully represent some first-person perspective?

Two probability distributions:

$\mathbf{P}_{\mathrm{3rd}}$: how **B** changes over time according to the prob. laws of **A**-world

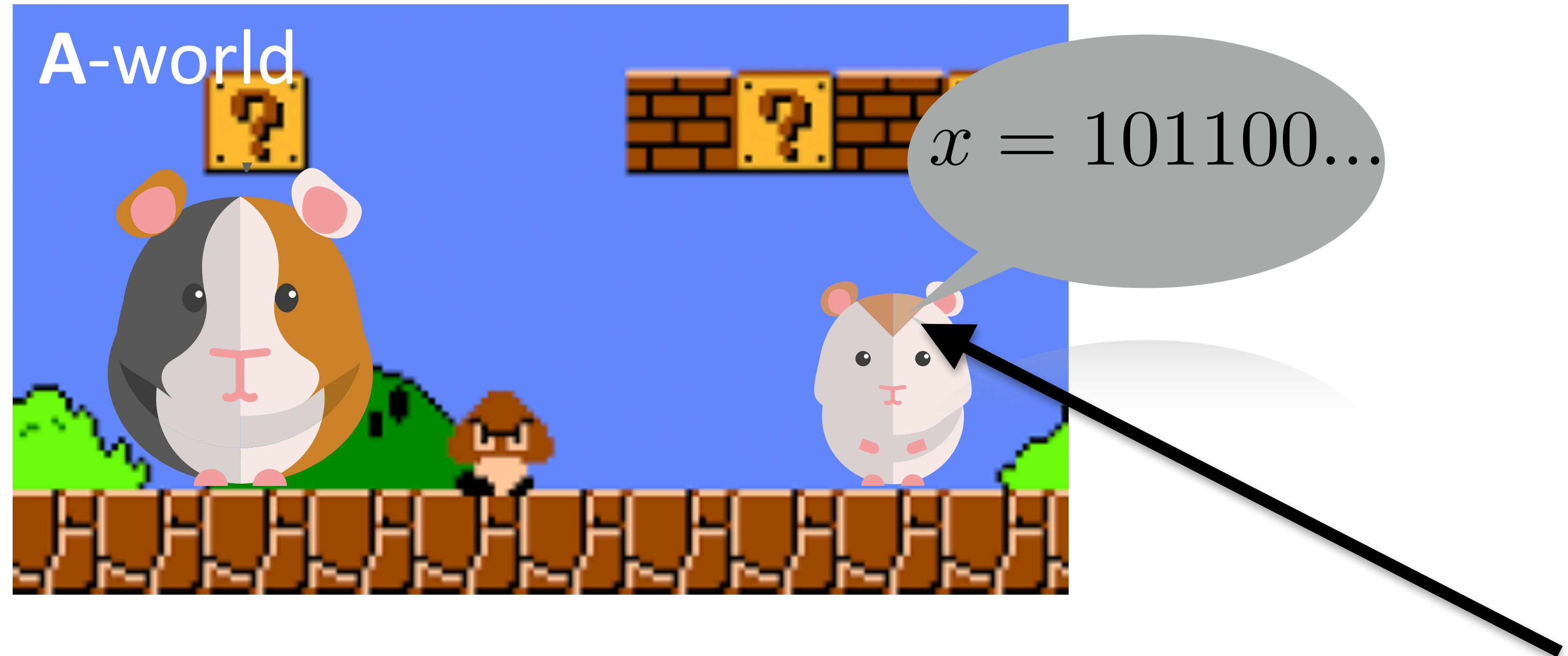
An emergent notion of **objective reality**



Alice

the guinea pig

... goes through patterns that look like...



Suppose in **A**-world, there is another bit-string valued random variable, **B**.

Does **B** faithfully represent some first-person perspective?

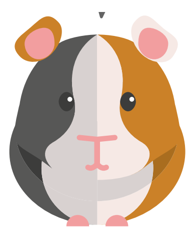
Two probability distributions:

$\mathbf{P}_{3\mathrm{rd}}$: how **B** changes over time according to the prob. laws of **A**-world

$\mathbf{P}_{1\mathrm{st}}$: the actual “first-person” chances of **B**ob’s self-pattern changes

$\mathbf{P}_{1\mathrm{st}} =$ universal probability **P**.

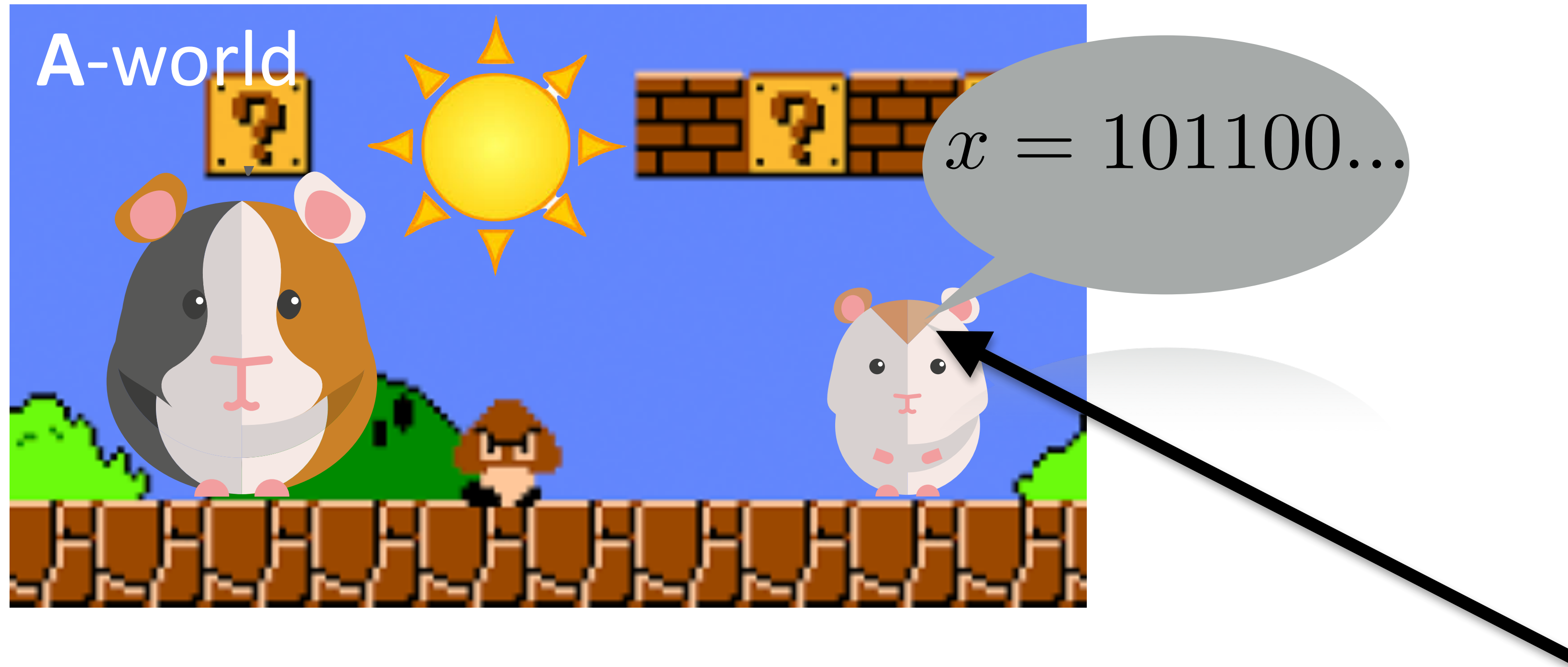
An emergent notion of **objective reality**



Alice

the guinea pig

... goes through patterns that look like...



Suppose in **A**-world, there is another bit-string valued random variable, **B**.

Does **B** faithfully represent some first-person perspective?

Two probability distributions:

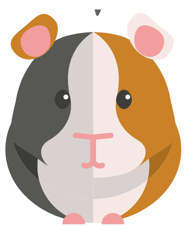
$\mathbf{P}_{\text{3rd}}$: how **B** changes over time according to the prob. laws of **A**-world

$\mathbf{P}_{\text{1st}}$: the actual “first-person” chances of **B**ob’s self-pattern changes

$\mathbf{P}_{\text{1st}} =$ universal probability **P**.

Example: If Alice has a 99% chance of seeing the sun rise tomorrow, and thus she has a 99% chance of seeing Bob see the sun rise tomorrow, will Bob’s actual chance of seeing the sun rise be 99%?

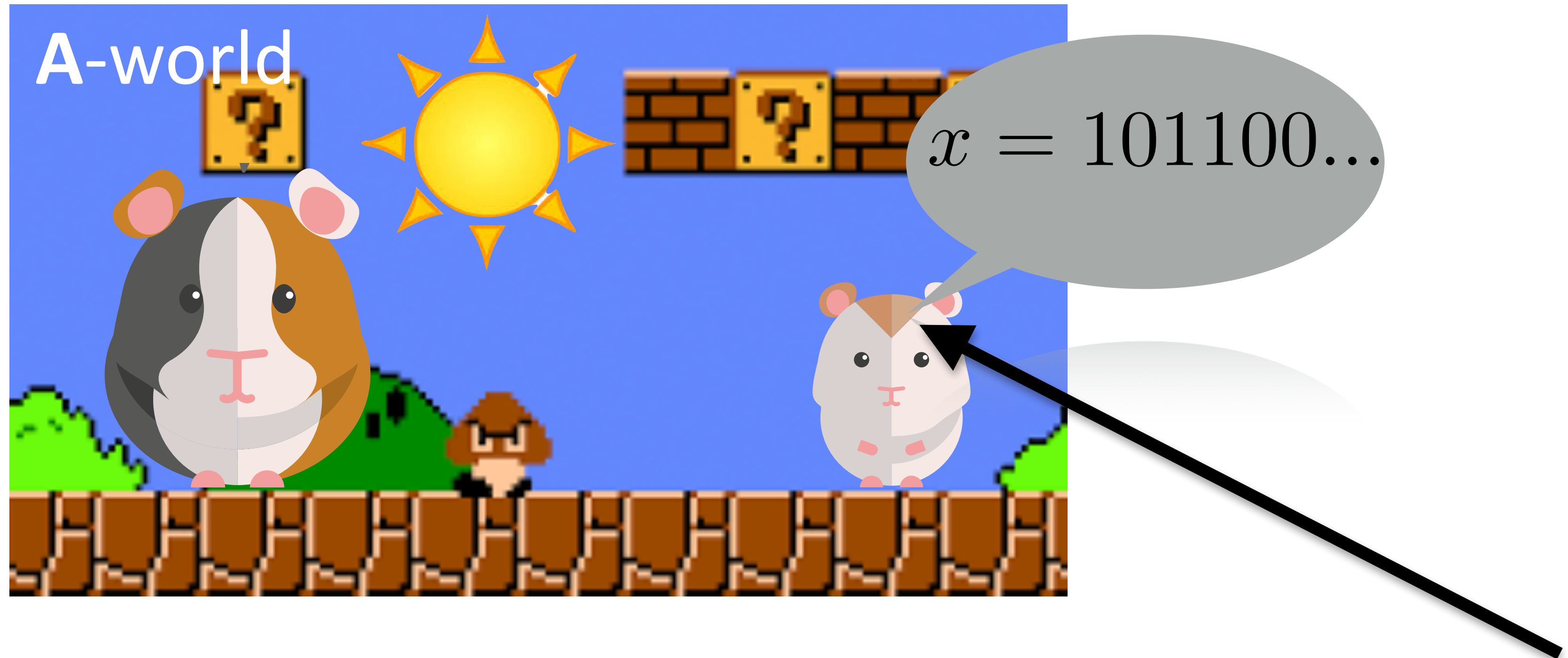
An emergent notion of **objective reality**



Alice

the guinea pig

… goes through patterns that look like…



Suppose in **A**-world, there is another bit-string valued random variable, **B**.

Does **B** faithfully represent some first-person perspective?

Two probability distributions:

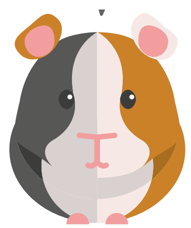
$\mathbf{P}_{3\mathrm{rd}}$: how **B** changes over time according to the prob. laws of **A**-world

$\mathbf{P}_{1\mathrm{st}}$: the actual "first-person" chances of **B**ob's self-pattern changes

$\mathbf{P}_{1\mathrm{st}} =$ universal probability **P**.

Example: If Alice has a 99% chance of seeing the sun rise tomorrow, and thus she has a 99% chance of seeing Bob see the sun rise tomorrow, $\mathbf{P}_{3\mathrm{rd}}$ will Bob's actual chance of seeing the sun rise be 99%? $\mathbf{P}_{1\mathrm{st}}$

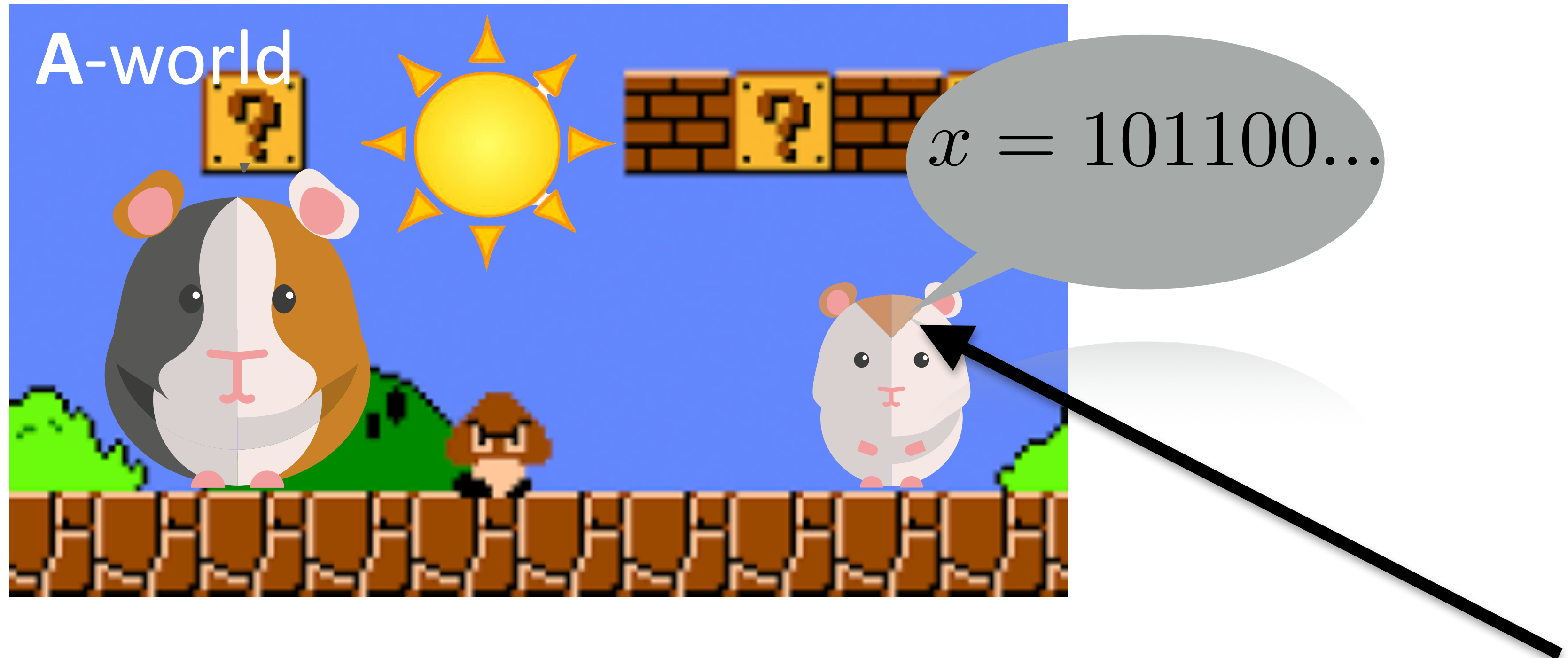
An emergent notion of **objective reality**



Alice

the guinea pig

… goes through patterns that look like…

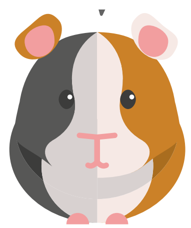


Suppose in **A**-world, there is another bit-string valued random variable, **B**.

Does **B** faithfully represent some first-person perspective?

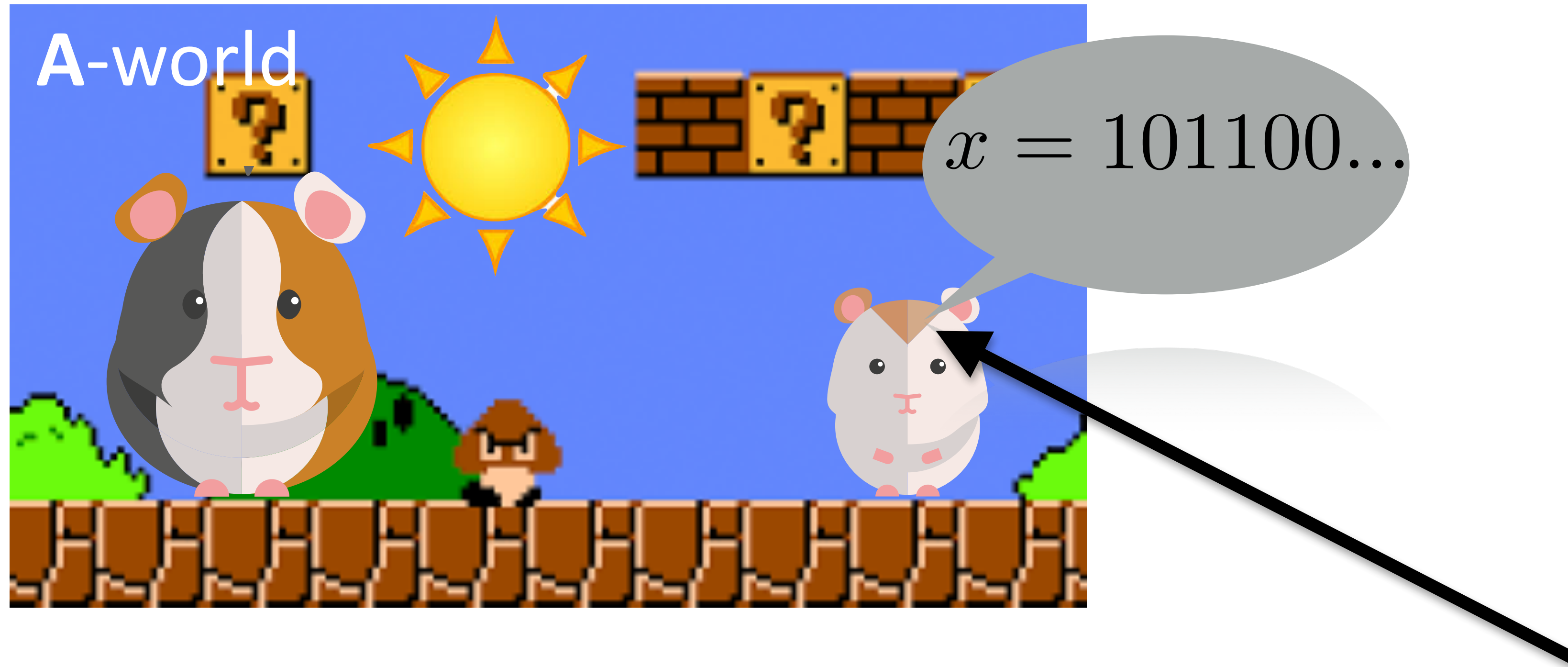
Example: If Alice has a 99% chance of seeing the sun rise tomorrow, and thus she has a 99% chance of seeing Bob see the sun rise tomorrow, $\mathbf{P}_{3\mathrm{rd}}$ will Bob's actual chance of seeing the sun rise be 99%? $\mathbf{P}_{1\mathrm{st}}$

An emergent notion of **objective reality**



Alice
the guinea pig

… goes through patterns that look like…



Suppose in **A**-world, there is another bit-string valued random variable, **B**.

Does **B** faithfully represent some first-person perspective?

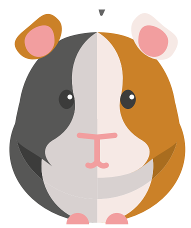
Theorem: As long as **B** keeps accumulating data without (much) forgetting,

$$|\mathbf{P}_{1\text{st}}(y|x_1,\ldots,x_n) - \mathbf{P}_{3\text{rd}}(y|x_1,\ldots,x_n)| \overset{n\to\infty}{\longrightarrow} 0,$$

so the answer is "yes": **A**-world = **B**-world.

Example: If Alice has a 99% chance of seeing the sun rise tomorrow, and thus she has a 99% chance of seeing Bob see the sun rise tomorrow, $\mathbf{P}_{3\text{rd}}$ will Bob's actual chance of seeing the sun rise be 99%? $\mathbf{P}_{1\text{st}}$

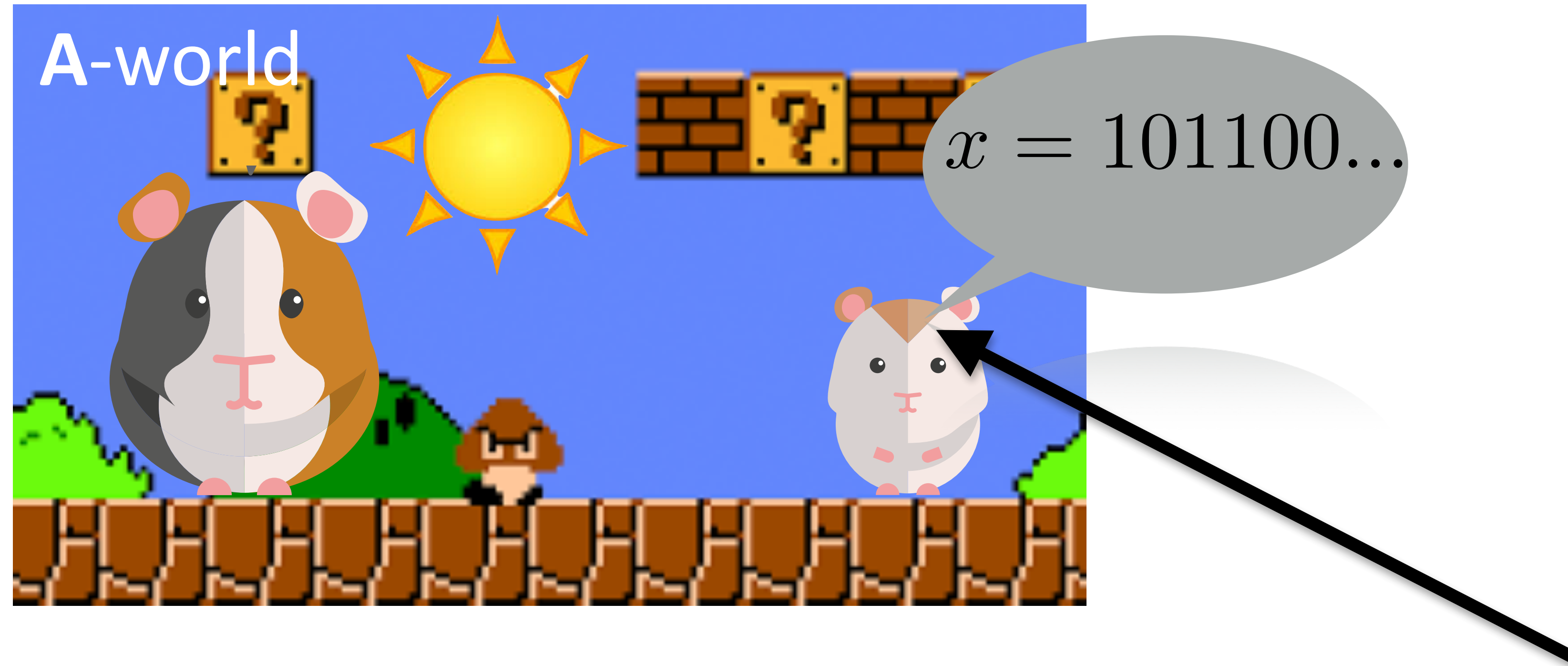
An emergent notion of **objective reality**



Alice

the guinea pig

… goes through patterns that look like…



Suppose in **A**-world, there is another bit-string valued random variable, **B**.

Does **B** faithfully represent some first-person perspective?

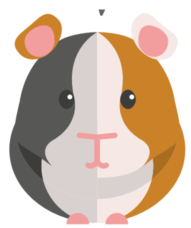
Theorem: As long as **B** keeps accumulating data without (much) forgetting,

$$|\mathbf{P}_{\mathrm{1st}}(y|x_1, \ldots, x_n) - \mathbf{P}_{\mathrm{3rd}}(y|x_1, \ldots, x_n)| \stackrel{n\to\infty}{\longrightarrow} 0,$$

so the answer is "yes": **A**-world = **B**-world.

"**Objective reality**" as a provable statistical phenomenon.

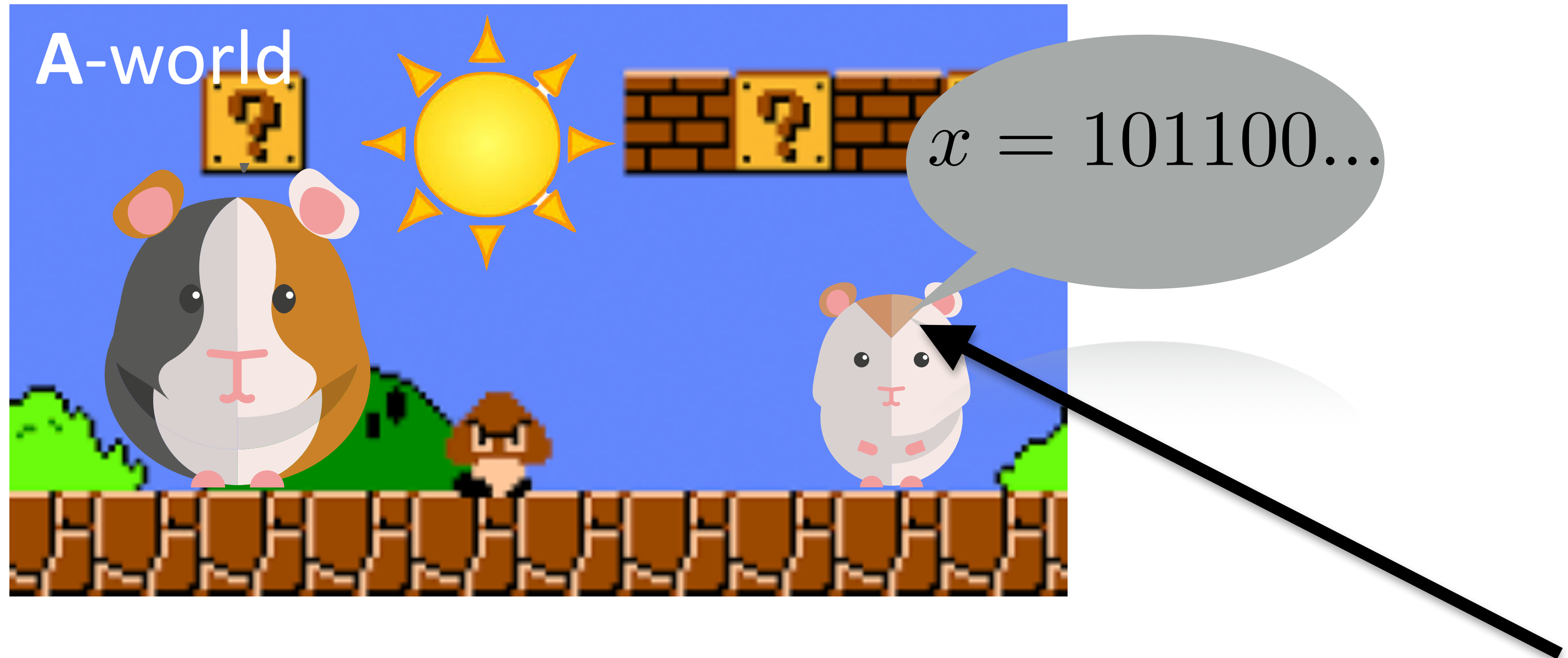
An emergent notion of **objective reality**



Alice

the guinea pig

... goes through patterns that look like...



Suppose in **A**-world, there is another bit-string valued random variable, **B**.

Does **B** faithfully represent some first-person perspective?

Theorem: As long as **B** keeps accumulating data without (much) forgetting,

$$|\mathbf{P}_{1\mathrm{st}}(y|x_1,\ldots,x_n) - \mathbf{P}_{3\mathrm{rd}}(y|x_1,\ldots,x_n)| \stackrel{n\to\infty}{\longrightarrow} 0,$$

so the answer is "yes": **A**-world = **B**-world.

"**Objective reality**" as a provable statistical phenomenon.

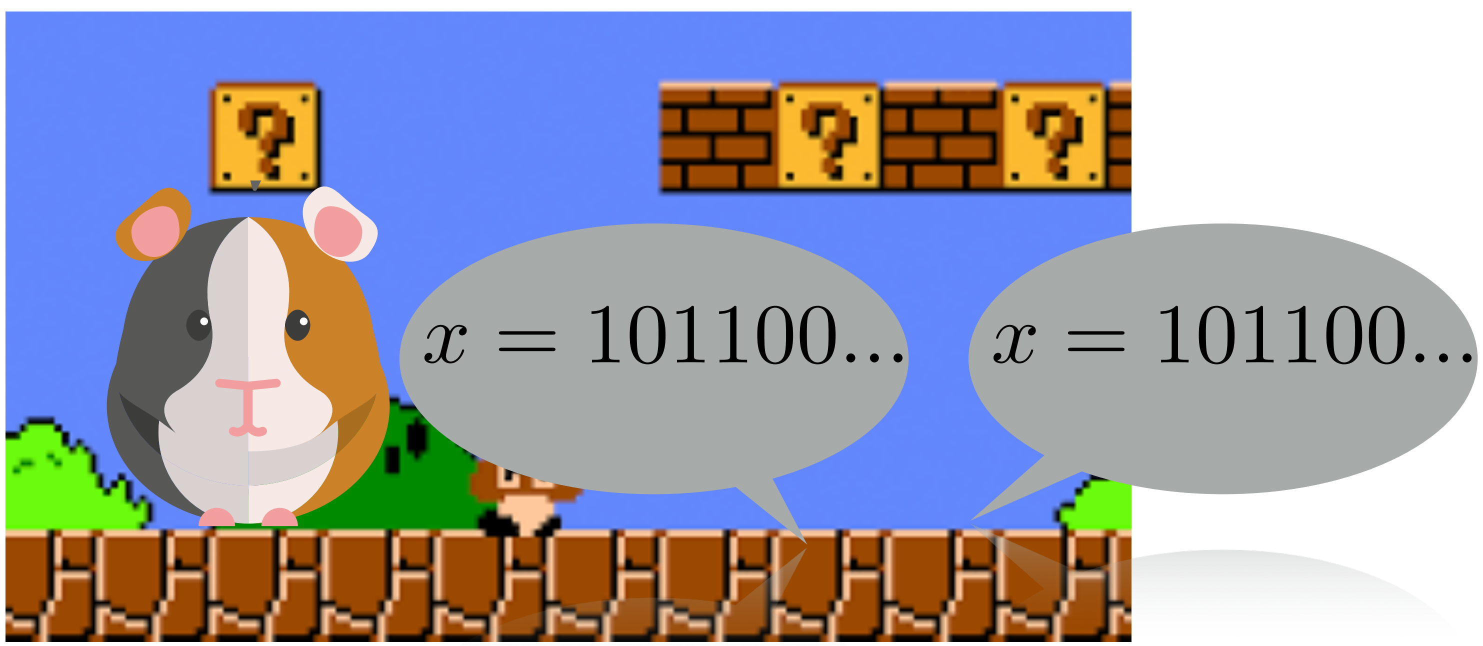
However, if **B** does not hold enough data, or forgets a lot (by accident), then $\mathbf{P}_{1\mathrm{st}} \not\approx \mathbf{P}_{3\mathrm{rd}}$ is possible. "Probabilistic zombie"

Probabilistic zombies

Probabilistic zombies

- Boring cases of $\mathbf{P}_{\text{1st}} \not\simeq \mathbf{P}_{\text{3rd}}$

  Self-patterns are just a bunch of information; need not be related to humans or guinea pigs.
  In A-world, Alice can simply copy a piece of information *x* to two places and **force the two instances to evolve differently**.



Then at least one of the two instances must have $\mathbf{P}_{\text{1st}}(y|x) \neq \mathbf{P}_{\text{3rd}}(y|x)$.

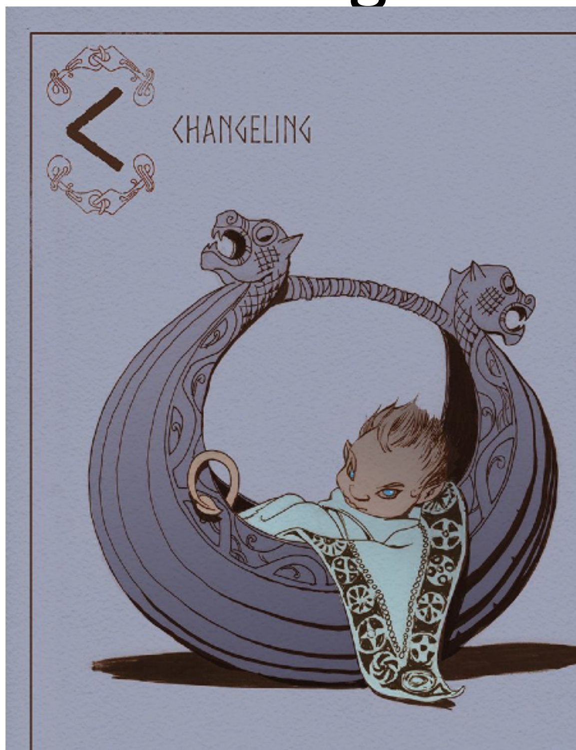
Probabilistic zombies

- Boring cases of $\mathbf{P}_{\text{1st}} \not\approx \mathbf{P}_{\text{3rd}}$

  Self-patterns are just a bunch of information; need not be related to humans or guinea pigs.
  In A-world, Alice can simply copy a piece of information *x* to two places and **force the two instances to evolve differently**.

- Disturbing cases of $\mathbf{P}_{\text{1st}} \not\approx \mathbf{P}_{\text{3rd}}$

  Alice runs a cellular automaton on her supercomputer for several years. Evolution kicks in, and after a long while, agents show up — including an agent called Bob who explores his cellular world and wonders about the meaning of it all. Then, suddenly, Alice intervenes in the simulation, say, by tuning its laws.
  Then, it is as if **"Bob's self leaks out of the simulation"** and becomes replaced by an unlikely changeling.



Pictures source: illustrator Thomas Denmark, https://thomden.artstation.com/projects/Ax5Pq

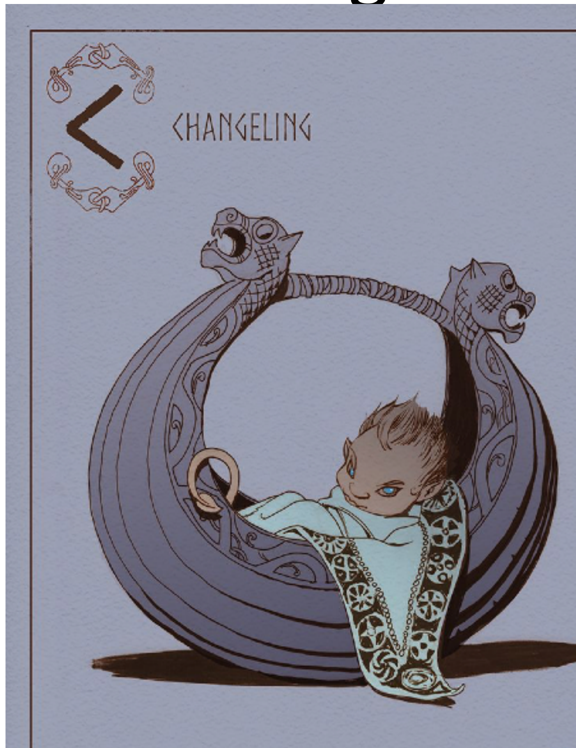
Probabilistic zombies

- Boring cases of $\mathbf{P}_{\text{1st}} \not\approx \mathbf{P}_{\text{3rd}}$

  Self-patterns are just a bunch of information; need not be related to humans or guinea pigs.
  In A-world, Alice can simply copy a piece of information *x* to two places and **force the two instances to evolve differently**.

- Disturbing cases of $\mathbf{P}_{\text{1st}} \not\approx \mathbf{P}_{\text{3rd}}$

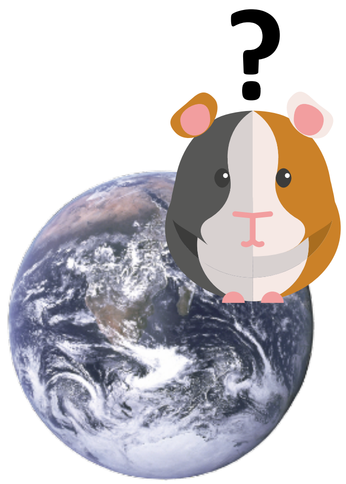
  Alice runs a cellular automaton on her supercomputer for several years. Evolution kicks in, and after a long while, agents show up — including an agent called Bob who explores his cellular world and wonders about the meaning of it all. Then, suddenly, Alice intervenes in the simulation, say, by tuning its laws.
  Then, it is as if **"Bob's self leaks out of the simulation"** and becomes replaced by an unlikely changeling.



Dissolves the puzzles…

Pictures source: illustrator Thomas Denmark, https://thomden.artstation.com/projects/Ax5Pq

Outline



1. **Conceptual puzzles**

   … that challenge the standard view.

2. **Sketch of an idealist (toy) theory**

   … "self" fundamental, external world emergent.

3. **Objective reality as a emergent approximation**

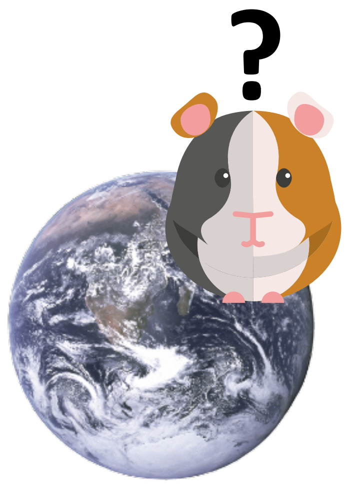
   … probabilistic zombies, and other surprises.

4. **Example: dissolution of the Boltzmann brain problem**

Outline

1. **Conceptual puzzles**

   … that challenge the standard view.



2. **Sketch of an idealist (toy) theory**

   … "self" fundamental, external world emergent.

3. **Objective reality as a emergent approximation**

   … probabilistic zombies, and other surprises.

4. **Example: dissolution of the Boltzmann brain problem**

Dissolution of the Boltzmann brain problem

Recall: Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

Dissolution of the Boltzmann brain problem

Recall: Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

next moment

Dissolution of the Boltzmann brain problem

Recall: Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

now

Dissolution of the Boltzmann brain problem

Recall: Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

now

Q: "Given what I see, and what I think I know, **am I the guinea pig on this planet or one of the BB quantum fluctuations?"**

Dissolution of the Boltzmann brain problem

Recall: Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

now

Q: "Given what I see, and what I think I know, **am I the guinea pig on this planet or one of the BB quantum fluctuations?"**

Standard-A: Count **how many BBs** there are, versus how many "standard guinea pigs" on planets. If there are far more BBs, then you are probably a BB and will soon disappear."

Dissolution of the Boltzmann brain problem

Recall: Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

now

Q: "Given what I see, and what I think I know, **am I the guinea pig on this planet or one of the BB quantum fluctuations?"**

~~**Standard-A:** Count **how many BBs** there are, versus how many "standard guinea pigs" on planets. If there are far more BBs, then you are probably a BB and will soon disappear."~~

Dissolution of the Boltzmann brain problem

Recall: Assume some ("combinatorially large") universe with a large number of "brains" with false memories fluctuating into existence.

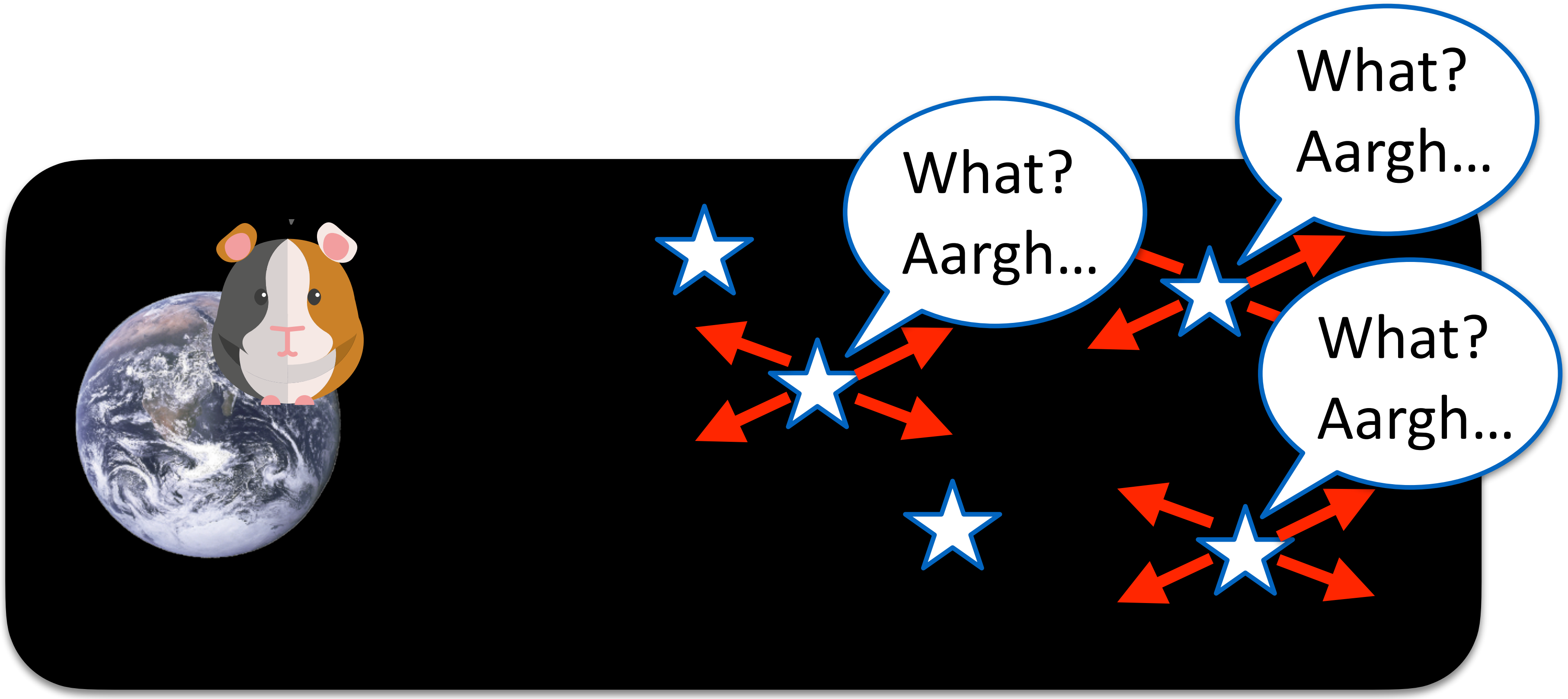
now

Q: "Given what I see, and what I think I know, **am I the guinea pig on this planet or one of the BB quantum fluctuations?"**

A: The question is meaningless. **You are your self-pattern.** This is **unembedded** structure that doesn't have a "position".
In some sense, you are all BBs and planet guinea pigs at once.

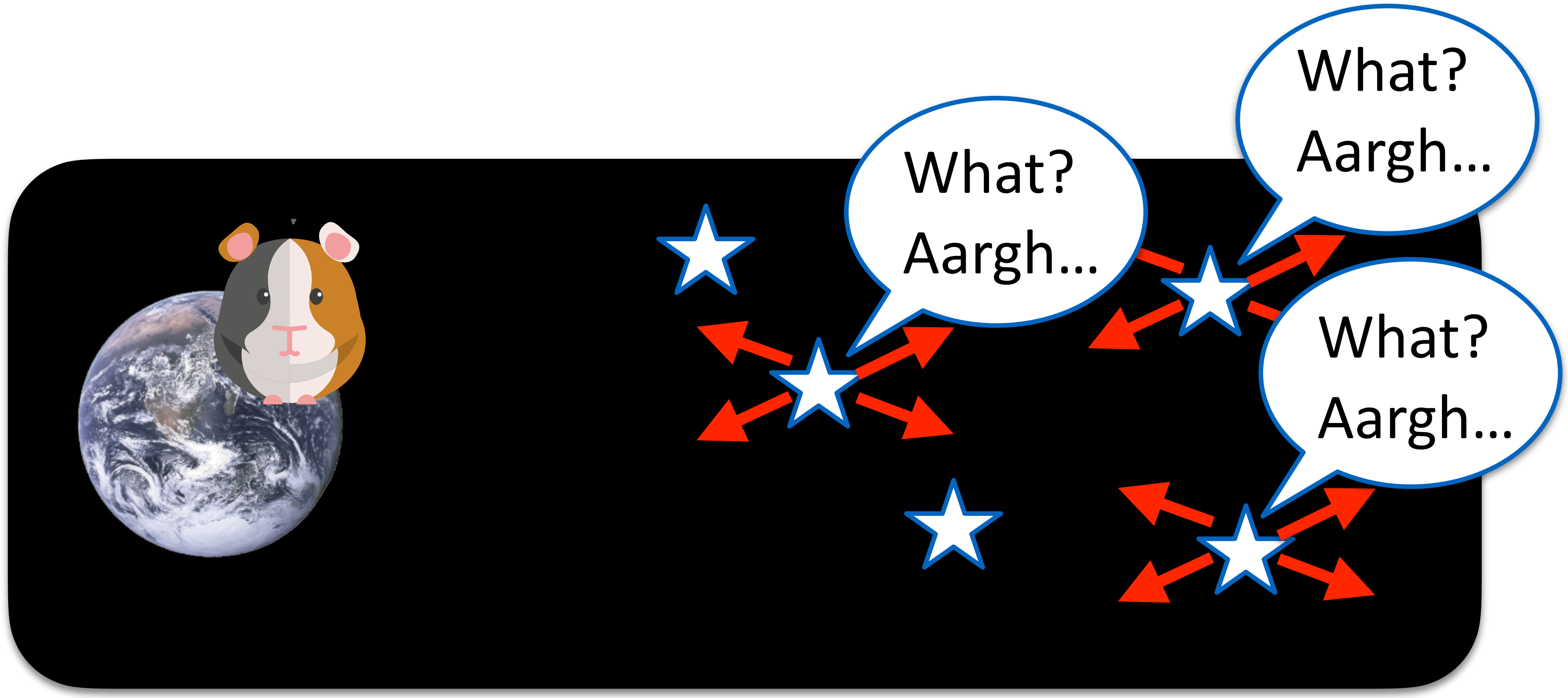
Dissolution of the Boltzmann brain problem

next?



Dissolution of the Boltzmann brain problem

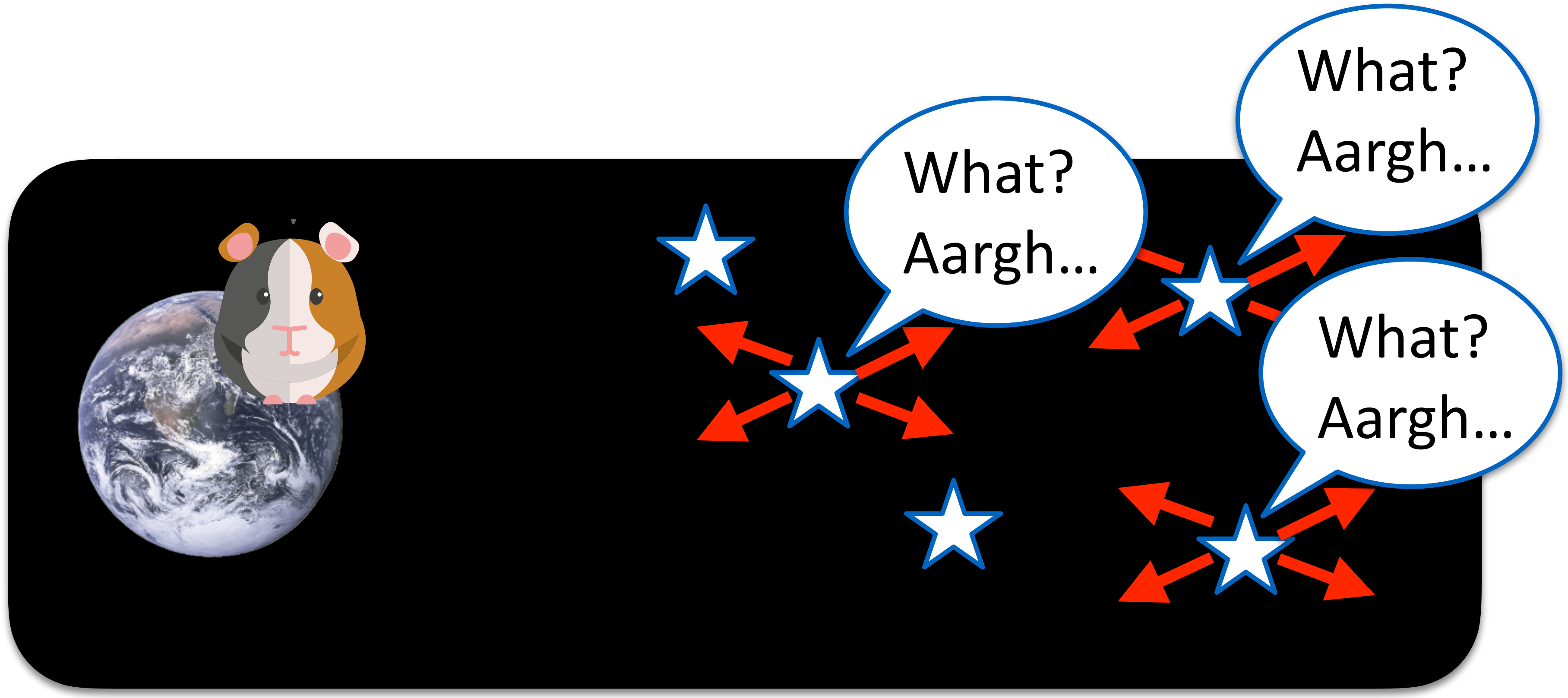
next?



Q: "Fair enough… but **what happens to me next?** Business as usual on Earth, or a strange BB experience?"

Dissolution of the Boltzmann brain problem

next?



Q: "Fair enough… but **what happens to me next?** Business as usual on Earth, or a strange BB experience?"

A: This **is** a meaningful question! You have to compare the universal probabilities $\mathbf{P}(y_{\mathrm{BB}}|x)$ versus $\mathbf{P}(y_{\mathrm{Earth}}|x)$.

Note: $\mathbf{P}(y|x)$ is larger if *y* is more **compressible**, given *x*.

Thus $\mathbf{P}(y_{\mathrm{Earth}}|x) \gg \mathbf{P}(y_{\mathrm{BB}}|x)$.

Business as usual will prevail, no matter how many BBs exist.

Conclusions

- Conceptual puzzles and Quantum Theory motivate **information-theoretic “idealist”** approach.

- Have shown an (incomplete toy) theory of this kind, based on **universal probability / algorithmic information theory**.

- Predictions: agents see a **simple, computable, probabilistic external world**; **objective reality** as an excellent approximation.

- Potential to **dissolve several relevant conceptual enigmas**, surprising new phenomena like “probabilistic zombies”.

M. P. Müller, Quantum **4**, 301 (2020)
Nontechnical paper in 2023 (hopefully).