Boltzmann Brains [2]

Introduction

Depending on our cosmological model, we may be a brain floating through an otherwise unordered fog of particles.

Descartes and Brains in Vats

Reality is a demon tricking me, so it's incoherent that I have to pay for half of the bill when I only ate a margarita pizza, and I also paid for the drinks earlier. My sensory evidence for the drinks payment is real, the demon wouldn't trick me like that.

Descartes (he didn't actually say this)

- Brain in a vat
  - Biological Objection
  - Difference in experience due to different stimuli
  - Putnam's Objection: Our words don't necessarily refer to what we think they do, therefore if a brain in a vat thinks/says 'I am a brain in a vat' it can only refer to brains and vats it has experienced in the simulation [3].
  - Putnam believes reference is achieved via 'causal connection'
  - Implicit argument from all living humans.
- Boltzmann: Where we're going, we won't need vats

Boltzmann

- Boltzmann
  - Classical to statistical physics.

$$S = k \log W$$

Second 'Law' of Thermodynamics:

$$\frac{\mathrm{d}S}{\mathrm{d}t} \geq 0$$

(i.e. S is always either increasing or constant)

Particles in a Partitioned Box, First State Space

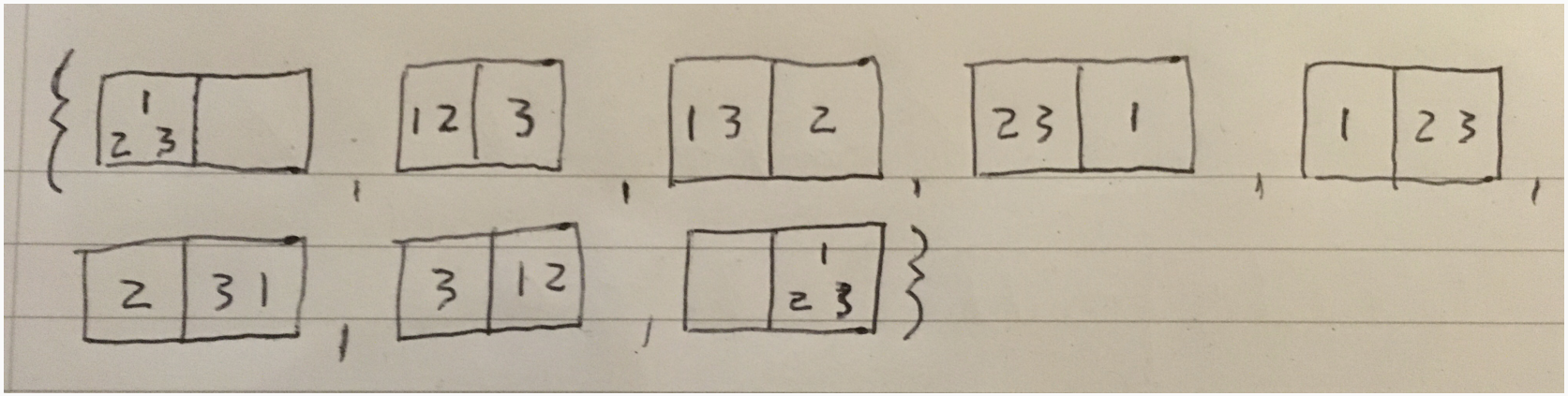


Figure 1: State space at the highest precision

Transitions

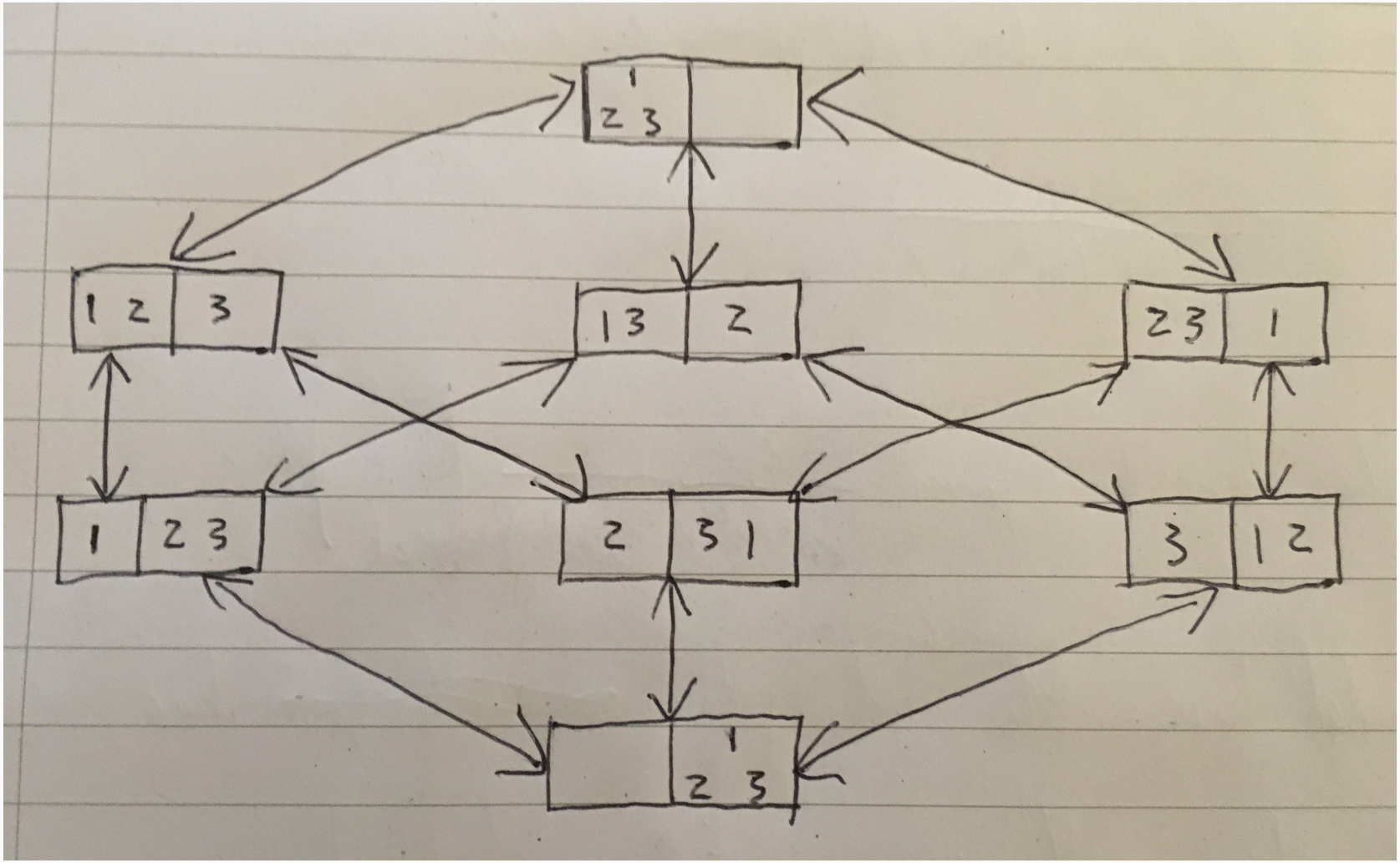


Figure 2: Transition Diagram

First Equilibrium

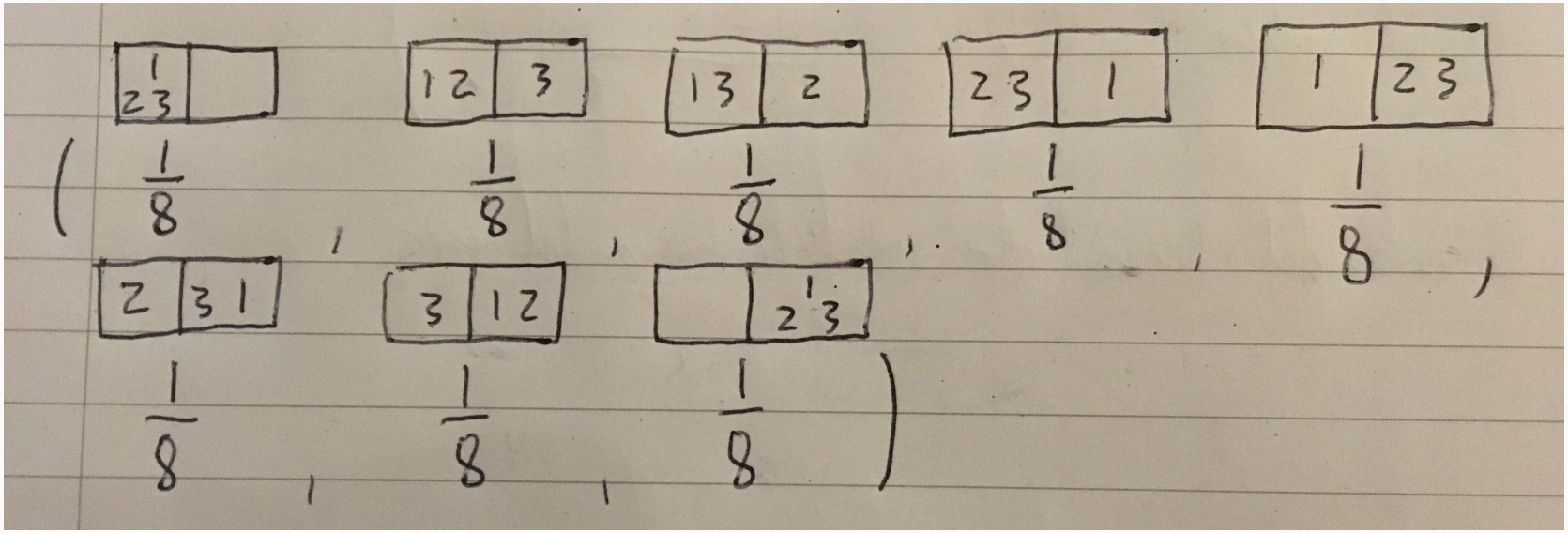


Figure 3: Equilibrium distribution at the highest precision

Lets get stupider

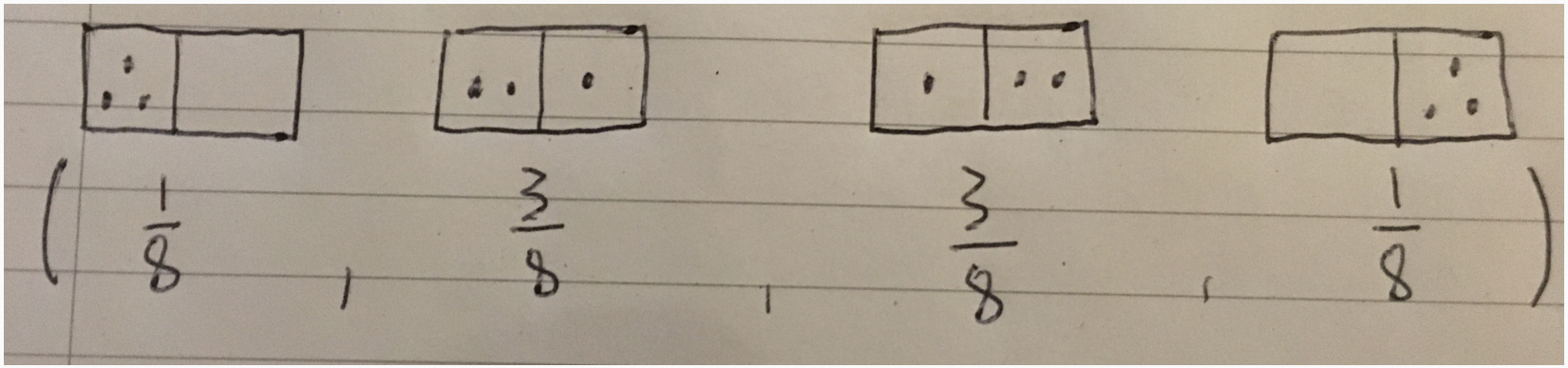
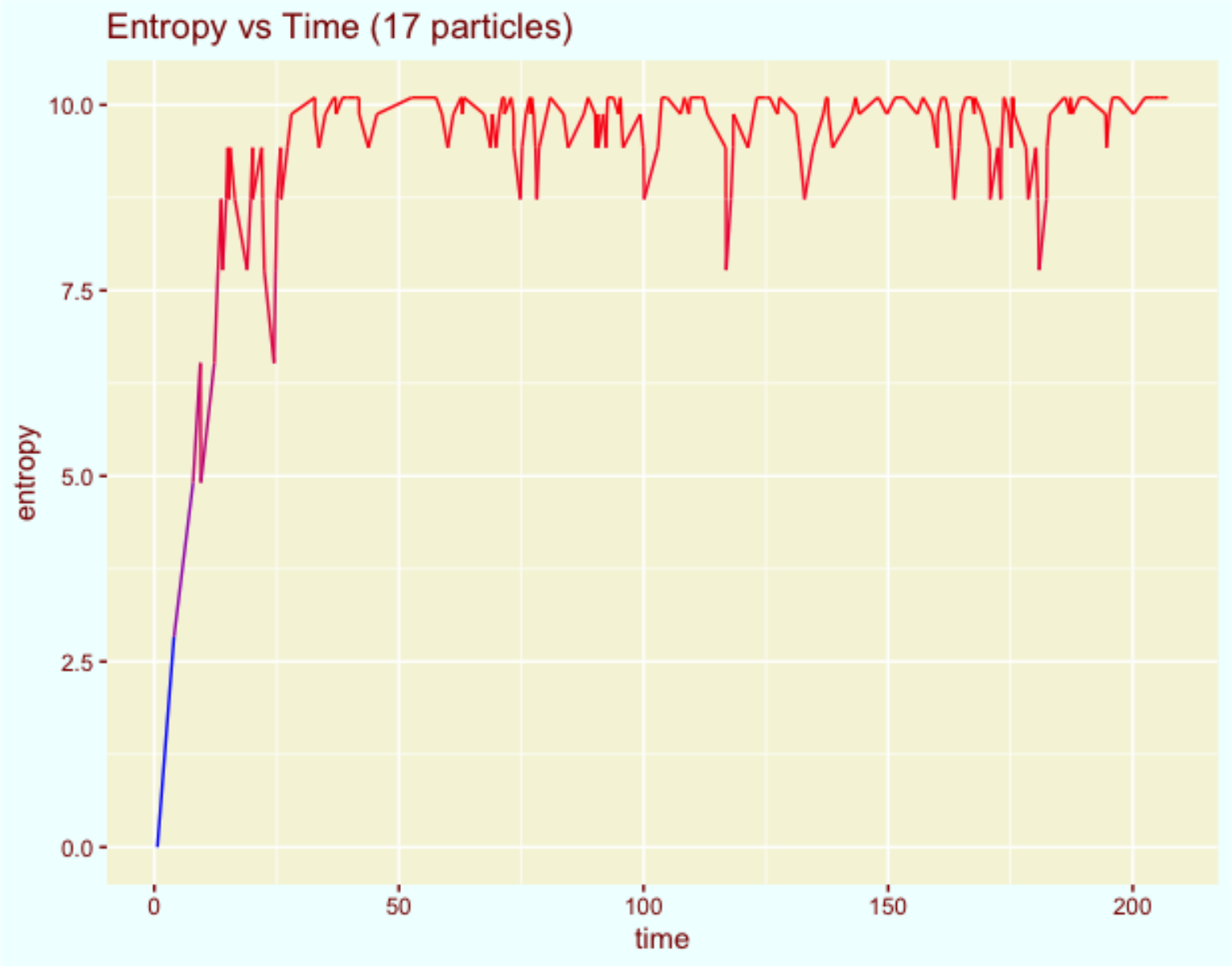


Figure 4: Equilibrium distribution at the new precision

Simulation


Entropy vs Time (17 particles)
entropy
10.0
7.5
5.0
2.5
0.0
0
50
100
150
200
time

Simulation II

There are 2 states with the lowest entropy (all particles in one chamber or the other)

There are

$$2 \times \binom{17}{8} = 48,620$$

states with the highest entropy ($2^{17} \approx 130,000$ states in total)

Poincaré Objects, Boltzmann replies

- Poincaré: The universe should recur.
  - First discussed by Poincaré (1890), formalized and proved using measure theory by Carathéodory (1919).
- Boltzmann: Ok, we might be near the start.
  - The past hypothesis

Descartes II: The Rise of Eddington

Eddington: Actually, all we need is for brains to exist. This is way more probable than a large region of low entropy.

Universes A and B

Quantifying Eddington's interjection: 2 Universes

1. Universe A: Young universe in which the 2nd law of thermodynamics holds.
2. Universe B: Universe old enough to exhibit thermal equilibrium and Poincare recurrence

2 types of observer:

1. Ordinary Observers (OOs): This is who we think we are (before hearing about Boltzmann brains/reading Descartes etc. etc.)
2. Boltzmann Brains (BBs)

Which observer are we?

$$N_{OO}(A) \approx N_{OO}(B) \leq 10^{124}$$

$$N_{BB}(A) \sim 0$$

$$N_{BB}(B) \sim e^{10^{122}}$$

So in universe A the probability of being an OO is basically 1, and in universe B the the probability of being a BB is basically 1 ($\approx \frac{e^{10^{122}}}{e^{10^{122}}+10^{124}}$)

Let's get Bayesian

Copernican Principle: For a given reference class of observers, we are equally likely to be any one of them.

Bayes' Theorem: If we have a set of theories $\{T_i\}$ that partition the space of all possible theories, and data D then

$$P(T_i|D) = \frac{P(D|T_i)}{P(D)} P(T_i)$$

We start with $T_A =$ The universe is A and $T_B =$ The universe is B. Our data is $D =$ Sense data (or local environment etc.) of OOs.

Naive calculation

$$P(D|T_A) \sim \frac{N_{OO}}{N_{OO} + N_{BB}(T_A)} = 1$$

$$P(D|T_B) \sim \frac{N_{OO}}{N_{OO} + N_{BB}(T_B)} = e^{-10^{122}}$$

$$\frac{P(T_A|D)}{P(T_B|D)} = \frac{\frac{P(D|T_A)}{P(D)} P(T_A)}{\frac{P(D|T_B)}{P(D)} P(T_B)} = \frac{P(D|T_A)}{P(D|T_B)} \frac{P(T_A)}{P(T_B)} = e^{10^{122}} \frac{P(T_A)}{P(T_B)}$$

Adjusting Priors: SIA

Self Indicating Assumption (SIA): we should form our priors (beliefs before data) as if we were chosen randomly from the set of possible observers given a theory

$$P(T_A) \propto N_{OO} + N_{BB}(T_A) \propto P(D|T_A)^{-1}$$

$$P(T_B) \propto N_{OO} + N_{BB}(T_B) \propto P(D|T_B)^{-1}$$

$$\frac{P(T_A|D)}{P(T_B|D)} = \frac{P(D|T_A)}{P(D|T_B)} \frac{P(T_A)}{P(T_B)} = \frac{P(D|T_A)}{P(D|T_B)} \frac{P(D|T_A)^{-1}}{P(D|T_B)^{-1}} = 1$$

Adjusting Priors: Presumptuous Philosopher

Bostrom: Weighting priors by number of observers introduces bias that makes us unscientific [1].

Adjusting Likelihoods: Jupiter

Hartle + Srednicki [4]: Our likelihoods shouldn't obey

$$P(D|T_i) \sim \frac{N_{OO}}{N_{OO} + N_{BB}(T_i)}$$

They imagine the following scenario:

$$P(\text{Life on Jupiter}) = \frac{1}{2}$$

$$P(\text{Intelligent population with } 10^{12} \text{ Jovians}|\text{Life on Jupiter}) = 1$$

So we have a half probability of only humans, a half probability of

$$\frac{\text{Jupiter's Population}}{\text{Total Population}} = \frac{10^{12}}{10^{12} + 10^{10}} = \frac{100}{100 + 1} \approx 99\%$$

Jovians

Adjusting Likelihoods: Jupiter II

$$
\begin{aligned}
P(\text{Life on Jupiter}|\text{We are Human}) &= \frac{P(\text{We are H}|\text{Life on J})}{P(\text{We are H})} P(\text{Life on J}) \\
&= P(\text{We are H}|\text{Life on J}) P(\text{Life on J}) \\
&= \frac{N_H}{N_H + N_J(\text{Life on J})} P(\text{Life on J}) \\
&= \frac{10^{10}}{10^{12} + 10^{10}} \frac{1}{2} \\
&= \frac{10^{10}}{10^{10}} \frac{1}{100 + 1} \frac{1}{2} \\
&= \frac{1}{202} < 0.01
\end{aligned}
\tag{1}
$$

Unreliable Data

- Nothing we do to our likelihood/priors (barring inclusion of extreme bias) can prevent the fact that D is best explained *for every* D by being a Boltzmann brain.
- Carroll: That's just no way to live [2].
- The argument that you are a Boltzmann brain is cognitively unstable:
- Man cannot do science from data alone: you need some philosophy/methodology/(???) which, under the Boltzmann brain scenario, was probably implanted in your brain by a random fluctuation.
- If ([our reasoning about the universe] $\implies$ 'we shouldn't trust our reasoning about the universe') we should reason differently.
- Carroll "It seems unreasonable to grant substantial credence to the prospect that we have no right to be granting substantial credence to anything"
- This doesn't discount the possibility of BBs.

Thoughts

- If our philosophy of science/(???) is due to a random fluctuation does that necessarily mean it's wrong?
  - How lucky would we have to be to get the "correct" philosophy of science as a BB?
- BBs may be infeasible: depends on philosophy of QM
- Making a prior probability 0 means the posterior probability is automatically 0, and if you were to use that posterior as a prior, the new posterior would be 0 etc. etc.
  - What this means is that literally no amount of data in support of the theory could give it a non-zero probability (bad imo)
- $p \implies$ not p is not a contradiction. If the implication is true, we would need p to be false, and here p is "our theory of physics"
  - This isn't actually what's going on with cognitive instability: we actually need "our theoretical/brain framework to support p is incidental" instead of not p
  - What difference does this make?
  - Can we find any parallels to other theories where this type of pattern of reasoning holds?

References i


N. Bostrom.
Anthropic Bias: Observation Selection Effects in Science and Philosophy.
Routledge, 2002.


S. M. Carroll.
Why boltzmann brains are bad, 2017.


H. Putnam.
Reason, Truth and History.
Cambridge University Press, 1981.


M. Srednicki and J. Hartle.
Science in a very large universe.
Physical Review D, 81(12), jun 2010.