Minds and Machines (Lecture 9): Scepticism and Computer Simulations

Prof. Ioannis Votsis

Philosophy Discipline
ioannis.votsis@nulondon.ac.uk
www.votsis.org

Introduction

A central question

- A venerable question in philosophy is the following epistemological question:

  What, if anything, can we know about the world?

- Those who either deny or withhold assent to the claim that we have some knowledge are known as sceptics.

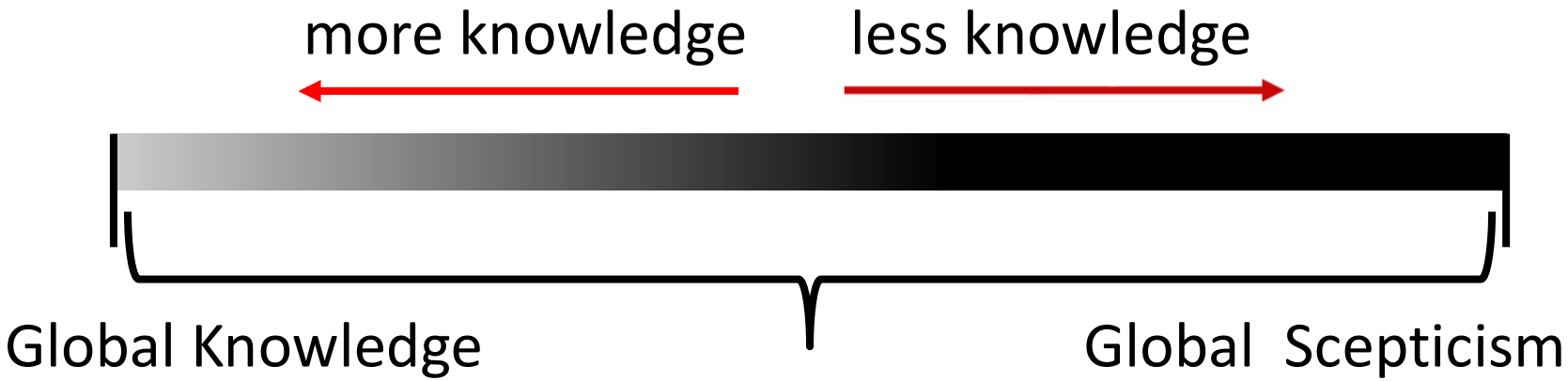
- Two branches of scepticism emerged in antiquity:

  * **Academic**: Knowledge of the external world is *impossible*.

  * **Pyrrhonian**: Undecidable if we can have such knowledge.

Scepticism: A spectrum?

- **Global scepticism**: We cannot know (or it's not decidable that we know) anything.
- **Local scepticism**: We cannot know (or it's not decidable that we know) some classes of things.
- A spectrum of views exists in between two extremes:



Global scepticism: An early example

- As is well known, Descartes puts forth two great sceptical scenarios: a dream scenario and an evil demon scenario.
- In both of them, he takes for granted that our senses *can be* (NB: not *are being*) deceived.
- So, is there anything that we can be sure of? Descartes' famous answer is: 'I think, therefore I am'.
- Two points deserve dwelling on here:

  (1) Some things cannot be doubted, e.g. rational thinking.

  (2) The most extreme version of global scepticism is foiled.

The modern debate

- Sceptical scenarios are nowadays prolific. A rough recipe for producing them involves:

  "imagining oneself to be in some possible world that is both vastly different from the actual world and at the same time absolutely indistinguishable (at least by us) from the actual world" (Klein in SEP entry on Skepticism).

- **Main point**: Our experience cannot pull us out of the thought that we may just be inhabiting the imagined world.

- This dovetails nicely with the underdetermination of theory by evidence thesis:

  The evidence is insufficient to determine which theory, among a number of rivals, is true.

The modern debate: Some scenarios

- A way to understand such scenarios is that they raise doubts about a default theory T_1 with one or more alternatives.

  T_1: I'm giving a Minds and Machines lecture.
  T_2: I'm a brain-in-a-vat being stimulated to believe T_1.

- The evidence (viz. experience) could be the product of:

  * the real world — **Non-sceptical scenario**
  * an imagination
  * a hallucination
  * a dream
  * an evil demon
  * a BIV world
  * a computer simulation

  — **Sceptical Scenarios**

Lecture plan

- In this lecture, we explore the scepticism that emerges out of cases of computer simulations.

- Plan:

  * What is a computer simulation?

  * What needs to be in place for computer simulations of minds like ours?

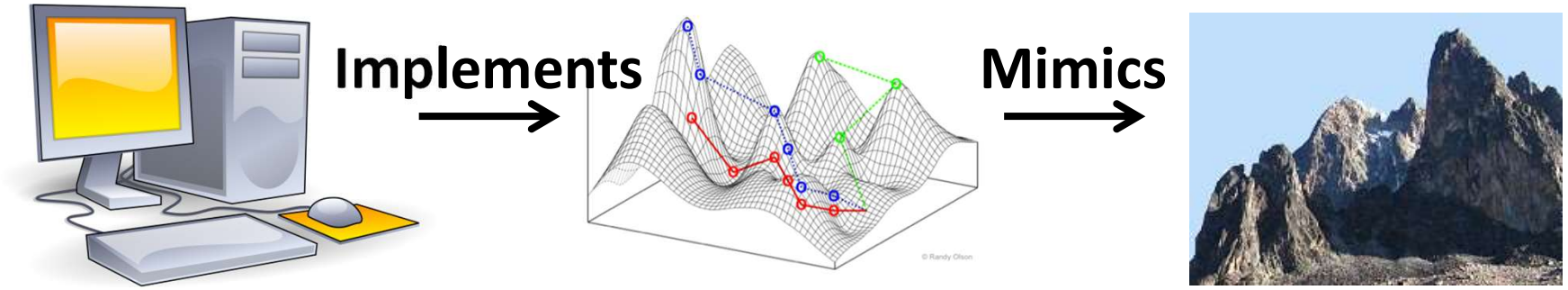
  * How likely is it that we are living in a computer simulation?

  * What can be said in reply to simulation arguments?

Computer Simulations

What is a computer simulation?

- A dynamical implementation of a math model (in a computer) that seeks to mimic some system.



- The mathematical model can be motivated by:

  * an underlying theory (e.g. the plate tectonics theory)
  * independent concerns (e.g. the Lotka-Volterra model)

Real vs. hypothetical systems

- It's worth noting that computer simulations are not exclusively concerned with the simulation of real systems.

- In some cases, the simulation of hypothetical systems is just as important:

  Example 1: A simulation can be used to probe the traits of a universe with radically different initial conditions to ours.

  Example 2: A simulation can be used to consider what would happen to goods prices if China were to join the EU.

- A hypothetical system is never completely different from a real one. Differences come in degrees.

Why simulate?

- A number of (oftentimes overlapping) reasons can be given in support of deploying computer simulations:

  * study a system
  * provide evidence for or against a theory
  * reduce costs
  * overcome limitations (ethical or technological)
  * offer a safe environment

  Example:
  "… in a simulation environment for a self-driving car, we could explore the limits of the car's performance, and any accidents give us more information. If the car crashes, we just hit the reset button" (Russell and Norvig 2020: 22.3.2).

A question of evidential import

- Philosophers of science are interested in whether the output of computer simulations has evidential import.

  "*How can scientists gain new knowledge by running computer simulations?* It seems puzzling that scientists can obtain knowledge about a real-world target system without actually observing it" (Beisbart 2012: 396).

- Most support the claim that computer simulations do have evidential import but disagree under what conditions.

- For example, Morrison (2009) thinks they're as good as real experiments. Barberousse et al. (2009) are less sanguine.

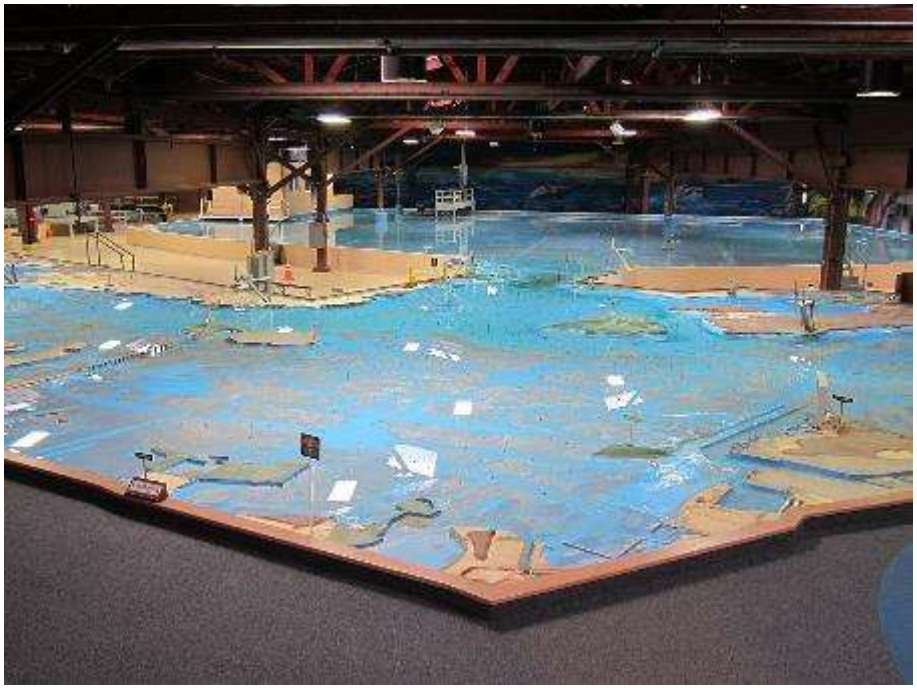
Simulations without computers

- A more generalised notion holds that a simulation:

  "… imitates one process by another process. In this [sense] the term 'process' refers solely to some object or system whose state changes in time" (Hartmann 1996: 83).

- **Example**: U.S. Army Corps of Engineers Bay Model

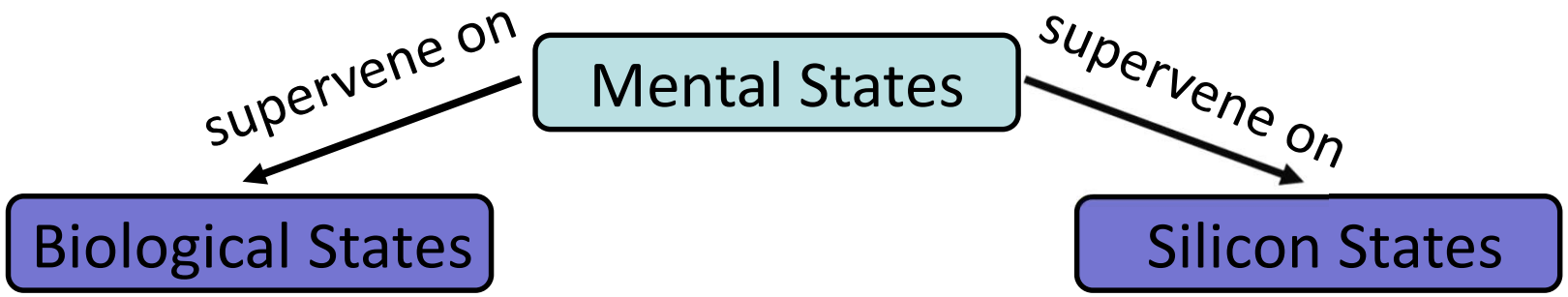
  * Hydraulic
  * Built in 1957
  * Purpose: to evaluate the Reber Plan (RB)
  * Result: RB rejected



Are we Living in a Computer Simulation?

Two assumptions

- **Substrate independence**: The substrate on which mental states supervene can manifest in various physical forms.



NB: The reason for this assumption is to allow for things other than biological brains to be conscious.

- **Post-human Capabilities**: Such civilisations will have enough computing power to run a vast # of ancestor simulations.

NB: We call such civilisations 'post-human'.

Simulations of minds: For what purpose?

- How likely is it that we are deceived into thinking that we live in the real world but are in fact living in a simulation?

- To answer this question, we first need to establish some initial plausibility regarding the purpose of such a simulation.

- On the assumption that a capacity to run such a computer simulation exists, why would anyone choose to do it?

- Given what we said in the previous section, the most obvious answer is for *scientific purposes*.

  Example: Predicting what would happen in a real world on the basis of a relevant class of simulated ones.

Bostrom's argument: A simplified version

1. "[E]normous amounts of computing power will be available in the future [of civilisations like ours]".

2. "[Such civilisations may] run detailed simulations of their forebears or of people like their forebears… [where the] simulated people are conscious".

3. "[T]hey could run a great many such simulations"

4. "[I]t could be the case that the vast majority of minds like ours [are] simulated"

---

5. "[I]f this were the case, we would be rational to think that we are likely to be among the simulated minds" (p. 243).

Bostrom's model

$$f_{sim} = \frac{f_p \bar{N} \bar{H}}{(f_p \bar{N} \bar{H}) + \bar{H}}$$

Where $\overline{N}$ is extremely large vis-à-vis $\overline{H}$, then $f_{sim} \approx 1$

f_{sim}: fraction of all observers (with human-like experiences) that live in simulations.

f_p: fraction of all civilisations with human-level technology that survive to become post-human.

$\overline{N}$: average # of ancestor simulations run by a post-human civ.

$\overline{H}$: average # of individuals who lived in a civilisation prior to its becoming post-human.

Bostrom's model: simplified

$$f_{sim} = \frac{number\ of\ simulated\ observers}{number\ of\ simulated\ observers + number\ of\ real\ observers}$$

From fractions to credences

- A credence is the degree of belief one has in a given claim.

  Example: Cr (SIM | evidence) is the credence we assign to the SIM claim, viz. that we are living in a simulation.

- How do we determine this credence? On the basis of the fraction of simulated observers.

  Cr (SIM | $f_{sim} = x$) = x

  NB: On the assumption that we have no other evidence!

- Since $f_{sim} \approx 1$ that means that our credence is close to 1. In short, we are almost certainly living in a computer sim.

Conditional version 1

1. If **it is rational to believe that we will have descendants who run lots of simulations of their ancestors**, **it is rational to believe that we are currently living in a simulation**.

2. **It is rational to believe that we will have descendants who run lots of simulations of their ancestors**.

---

3. **It is rational to believe that we are currently living in a simulation**.

Modus Ponens Schema:

1. If P then Q
2. P

---

3. Q

Conditional version 2

1. If **it is rational to believe that we will have descendants who run lots of simulations of their ancestors**, **it is rational to believe that we are currently living in a simulation**.

2. **It is *not* rational to believe that we are currently living in a computer simulation**.

---

3. **It is *not* rational to believe that we will have descendants who run lots of simulations of their ancestors**.

Modus Tollens Schema:

1. If P then Q
2. ¬Q

---

3. ¬P

Creating simulations → likely we are living in one

- Suppose humanity creates ancestor simulations at some point. This would be evidence against $f_p \approx 0$ and $f_l \approx 0$.

- That's because, on the assumption that what happens to us is not special, we would have to assert that:

  * almost all civilisations like ours reach a post-human stage
  * almost all post-human civilisations create ancestor sims.

- We would have strong evidence that $f_{sim} \approx 1$ and credence in the living-in-a-simulation hypothesis would be very high.

  NB: Bostrom argues that such simulations (and the civs within them) would be stacked and reality multi-layered.

Some objections

- **Framing assumptions implausible**: For example, it may be argued that conscious experience is tied to human biology.

- **Anthropomorphic**: Most other observers in the universe may be such that they don't have human-like experiences.

  So long as the few, if any, that do have them produce fewer, if any, ancestor simulations, Cr ($SIM_H \mid f_{simH} \approx 0$) ≈ 0.

- **Scientific purpose implausible**: If a civ had a theory of everything, it wouldn't need to simulate ancestors.

- **Extra evidence**: If civs only simulate interesting epochs, then we are not likely to be simulated (Weatherson 2003).

A superintelligence twist

- Arguably, a civ w/the capacity to run such simulations is also likely to be able to build an artificial superintelligence (ASI).

  NB: Assuming, of course, that ASIs can be constructed.

- Arguably, such a task would be given to the ASI as it could more easily outwit attempts to detect the deception.

- So, if we are a living in a computer simulation, we are likely to have been created by an ASI.

- Quite a few people question the reality of ASIs. Ironically, it might be WE who are less real than the ASIs!

The End