POS Lab: Simulation Hypothesis

Dr. Yves J. Hilpisch[1]

December 11, 2025

Purpose and Connection

This lab applies ideas from *Introduction to the Philosophy of Science: Concepts, Practice, and Case Studies* [1] to Nick Bostrom's simulation argument [2] and later developments by David Chalmers [3], Robin Hanson [4], David Wolpert [7], and critics such as Sabine Hossenfelder [8]. Use it as a guide to interrogate the Simulation Hypothesis through measurement, evidence, computation, and reflexivity.

[1]Get in touch: `https://linktr.ee/dyjh`. Web page `https://hilpisch.com`. Research, structuring, drafting, and visualizations were assisted by GPT 5.1 as a co-writing tool under human direction.

Contents

1 Scenario: A Simulation Hypothesis Task Force

Imagine you are part of a small interdisciplinary task force—philosophers of science, computer scientists, and physicists—asked to produce a research dossier on the Simulation Hypothesis for a funding body.

High-Level Narrative

- Bostrom's trilemma [2] frames the stakes: extinction before posthumanity, disinterest in ancestor simulations, or high odds that we are simulated.
- Experimentalists want concrete tests; theorists worry about testability and anthropic reasoning.
- Stakeholders care about ethical, epistemic, and governance implications even if the hypothesis remains unproven.

Your task in this lab is to structure that dossier using the measurement, evidence, causality, complexity, and reflexivity tools from the main book [1].

From Lab to Life

Keep a live use case in mind: a university centre debating whether to fund experimental probes (for example cosmic-ray anisotropy checks), a think tank writing a public explainer, or an ethics board considering how the hypothesis affects decision-making. Each exercise should feed a memo convincing sceptical reviewers that your recommendations are both philosophically rigorous and practically grounded.

2 Framing the Argument and Priors

Relate this section to Chapter 2 on evidence and Chapter 4 on hypothesis appraisal.

Step 1: Restate the Trilemma

Write down Bostrom's three propositions in your own words [2]. For each, list at least two auxiliary assumptions (about technological limits, civilisation lifetimes, or motivations) that must hold for the proposition to be plausible. Mark which assumptions are most brittle.

Step 2: Priors and Anthropics

Sketch a Bayesian update:

- start with a prior over each proposition,
- note what kinds of evidence could shift these priors (for example successful mind uploads, observed limits to computation, or signs of cosmic resource use),
- write one paragraph on how anthropic reasoning enters the update and where it risks circularity, following the cautionary stance from Chapter 2 of the main text [1].

Substrate Independence

Explain why the argument relies on substrate independence—that consciousness could in principle run on computational substrates [2, 3]. Identify one philosophical reason to doubt this and one empirical line of work (for example neuro-computation mappings) that could strengthen it.

3 Evidence, Testability, and Falsifiability

Relate this section to Chapter 3 on theories and tests and Chapter 5 on measurement.

Step 3: Observable Signatures

List plausible observational signatures often discussed in the literature:

- computational artefacts such as lattice-like limits on high-energy particle trajectories [8],
- information-theoretic constraints like Vopson's proposed "second law of infodynamics" [6],
- resource-usage footprints suggested by Lloyd's "universe as quantum computer" analogy [5].

For each, classify whether it is a genuine test of the Simulation Hypothesis or merely a speculative analogy.

Step 4: A Testability Matrix

Create a small table (even hand-drawn) with rows for candidate tests and columns for *measurement*, *model*, *expected deviation*, and *interpretation*. Rate each test on: scientific legitimacy, vulnerability to underdetermination, and risk of post-hoc rationalisation.

For a bird's-eye reminder of the trilemma logic that frames these tests, Figure 1 highlights how extinction risk, simulator motivation, and anthropic counting interact.

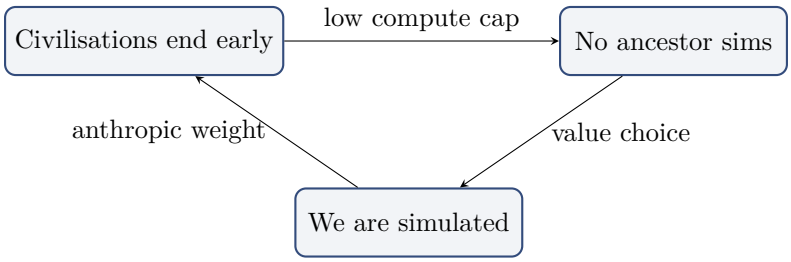

Figure 1: A stylised view of Bostrom's trilemma [2]: each prong depends on auxiliary claims about survival, motivation, and anthropic counting.

4 Computation, Complexity, and Feasibility

Relate this section to Chapter 9 on complexity and Chapter 12 on computation in the main book [1].

Step 5: Budgeting the Simulation

Draft a back-of-the-envelope compute budget:

- estimate the state space or information content needed to simulate a human brain, a city, or a planetary surface,
- contrast classical vs. quantum computing assumptions (drawing on [5]),
- note how coarse-graining, compression, or lazy rendering ("render on observation") change the budget and invite comparisons to video-game engines.

Step 6: Self-Simulation and Regress

Summarise Wolpert's result that self-simulation is mathematically possible under physical Church–Turing assumptions [7]. Ask: does this avoid infinite regress or merely shift the burden to resource constraints? Connect this to the complexity warnings in Chapter 9 about models that outstrip available computation.

Compression vs. Hidden Complexity

Frank Wilczek has argued that physics contains "hidden complexity" that would be unnecessary in an efficient simulation [9]. Contrast that view with pro-simulation arguments that prioritise compression and error bounds [6]. Decide which side you find more plausible and why.

5 Semantics, Meaning, and Consciousness

Relate this section to Chapter 6 on causality and Chapter 11 on models and semantics.

Step 7: Virtual Objects as Real

Using Chalmers' "The Matrix as Metaphysics" [3], write a short explanation of how simulated tables, trees, and agents can still count as real within the simulation. Identify one implication for measurement: when you observe a lab instrument, what exactly are you measuring if both you and the instrument are computational structures?

Step 8: Consciousness and Substrate

List two reasons to think consciousness could be substrate-independent and two reasons to doubt it. Map how each reason would affect the Simulation Hypothesis: does it undermine the entire argument or merely shift probabilities?

6 Reflexivity, Ethics, and Decision-Making

Relate this section to Chapter 13 on reflexivity and Chapter 8 on decision under uncertainty.

Step 9: Policy and Behaviour in a Simulated World

Following Hanson [4], propose one behavioural policy that might make sense if you suspected simulation (for example being "interesting" to simulators) and one policy that would be irrational even if the hypothesis were true. Explain how reflexive effects could arise if many agents adopted the same policy.

Step 10: Governance Checklist

Draft a checklist (5–7 items) for research governance:

- criteria for distinguishing scientific tests from speculative narratives,
- guidelines for public communication to avoid pseudoscientific claims [8],
- ethical guardrails for experiments that might affect the simulated environment or its observers.

7 Putting It Together: A Mini Research Dossier

As a final exercise, write a two-page dossier that:

- restates the core hypotheses and priors,
- summarises proposed tests and how they map to observables,
- reports compute and complexity considerations,
- addresses semantic and consciousness assumptions,
- discusses reflexive and ethical implications,
- lists concrete next experiments with timelines and decision criteria.

Try in 60 Seconds

If time is short:

- jot one empirical signature you would actually fund and one you would cut as unfalsifiable,
- state one way the Simulation Hypothesis sharpens your thinking about measurement and one way it risks distracting from testable science.

References

[1] Y. J. Hilpisch. *Introduction to the Philosophy of Science: Concepts, Practice, and Case Studies*. 2025. Available at hilpisch.com.

[2] N. Bostrom. Are you living in a computer simulation? *Philosophical Quarterly*, 53(211):243–255, 2003. Available at simulation-argument.com.

[3] D. Chalmers. The Matrix as metaphysics. 2005. Available at consc.net.

[4] R. Hanson. How to live in a simulation. *Journal of Evolution and Technology*, 7, 2001. Available at jetpress.org.

[5] S. Lloyd. How to test if we're living in a computer simulation. University of Portsmouth blog post, 2022. Available at port.ac.uk.

[6] M. Vopson. The second law of infodynamics and its implications. *AIP Advances*, 13(10):105308, 2023. Available at pubs.aip.org.

[7] D. H. Wolpert. Implications of computer science theory for the simulation hypothesis. 2024. Available at arxiv.org.

[8] S. Hossenfelder. The simulation hypothesis is pseudoscience. 2021. Available at backreaction.blogspot.com.

[9] Anonymous. Simulation hypothesis. 2025. Available at en.wikipedia.org.