


A Theodicy for Artificial Universes: Moral Considerations on Simulation Hypotheses

Stefano Gualeni, Institute of Digital Games, University of Malta, Malta

 <https://orcid.org/0000-0003-4873-4751>

ABSTRACT

“Simulation hypotheses” are imaginative scenarios that are typically employed in philosophy to speculate on how likely it is that one is currently living within a simulated universe as well as on the possibility for ever discerning whether one does in fact inhabit one. These philosophical questions in particular overshadowed other aspects and potential uses of simulation hypotheses, some of which are foregrounded in this article. More specifically, “A Theodicy for Artificial Universes” focuses on the moral implications of simulation hypotheses with the objective of speculatively answering questions concerning computer simulations such as: If one is indeed living in a computer simulation, what might be its purpose? What aspirations and values could be inferentially attributed to its alleged creators? And would living in a simulated universe affect the value and meaning one attributes to the existence?

KEYWORDS

Artificial Agents, Existential Meaning, Existentialism, Morality, Posthuman Morality, Posthuman Morality Hypothesis, Simulation Argument, Simulation Hypothesis, Theodicy, Utilitarianism

INTRODUCTION

Imagine a large vat on the table of a futuristic laboratory. Inside the vat, a disembodied brain floats in some kind of liquid. The scientists running the laboratory use advanced computer technology to stimulate the brain in the vat with input and sensations that are indistinguishable from those that regular human bodies experience in their relationship with the actual world. In this hypothetical setup, the laboratory’s technology also feeds the brain’s outputs back into the computer, giving the brain the possibility to interact with the environment it perceives. At that point, the brain is effectively inhabiting a persistent, interconnected whole: a world¹. In this hypothetical scenario, the brain floating in the vat is connected with what is commonly referred to as a simulation: a procedural model – often run on computers – that imitates the behaviors of a physical system (see Bostrom, 2003; Salen & Zimmerman, 2003, 423; Chalmers, 2005).

Whether or not those imaginative scenarios explicitly rely on the use of digital technologies, speculative premises similar to the one that was just outlined are common throughout the history of Western thought. The Socratic dialogues and the texts of the skeptics feature questions, allegories, and ideas that can be considered particularly obvious examples of this recurrence. Within our tradition of thought, these propositions are often referred to as the ‘brain in a vat hypothesis’ (or the ‘evil genius

DOI: 10.4018/IJT.2021010102

Copyright © 2021, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

hypothesis', after René Descartes's infamous argument in the *Meditations*). Comparable ideas also emerged in non-Western cultural contexts such as Chinese Taoism or Vedic literature, where they similarly function as conceptual tools to help the reader maintain a degree of suspicion towards the emotions and sensations that they experience in their daily lives as embodied beings. They are, to cite Chalmers's (2005) words, 'philosophical fables' that prompt us question what we mean by 'reality' and what qualifies as a real experience. They invite us to consider whether the world of sensations and relationships that we experience on an everyday basis could be an artifice or a mere illusion.

In recent years, a particular set of speculative scenarios that are to a degree similar to those mentioned above received sustained attention both in academia and in popular culture. I am referring to a group of hypothetical situations that are commonly grouped under the umbrella term 'the simulation hypothesis' (SH). Differently from 'brain in a vat' kinds of hypotheses, the SH does not predicate that one's brain – or the entirety of one's body, as is the case in the movie *The Matrix* – exists somewhere in base reality². It does not presuppose that our perceptual and cognitive equipment is being enthralled and deceived by the simulative capabilities of a computer (or the magical ones of a demon). Instead, the SH proposes an imaginative scenario in which we are artificial beings who were created – and presently exist – within a computer simulation. In the SH, in other words, no part of me is predicated to be existing or having ever existed in base reality (with the exclusion, perhaps, of the figments of computer code that correspond to the properties my being and my mental states)³.

Like 'brain in a vat' kinds of hypotheses, the various versions of the SH are also used to raise doubts concerning the artificiality of our experience. They are similarly employed in philosophy to gauge the likelihood of our being currently living in a simulated universe as well as our capability of ever discerning whether we do in fact inhabit one. This dominant philosophical use sidelined other aspects and potential applications of simulation hypotheses, some of which are foregrounded in the present article. I am talking, for example, about reflections concerning the technological and computational requirements that would be needed to run the simulation in question, about the kinds of values and aspirations that could have shaped and guided the design of that simulation, or about the ethical responsibilities that the creators of the simulation potentially have towards the artificial beings inhabiting it.

Notable existing work on those arguably secondary aspects include the notorious article "Are You Living in a Computer Simulation?" by Nick Bostrom (2003, which will be introduced and discussed in the next section), "Theological Implications of the Simulation Argument" by Eric Steinhart (2010), and "Natural Evil and the Simulation Hypothesis" by David Kyle Johnson (2011).

THE 'POSTHUMAN MORALITY HYPOTHESIS' (PMH)

In "Are You Living in a Computer Simulation?", Bostrom argues that at least one of the three following propositions must be true:

1. The human species is very likely to go extinct before reaching a stage of technological maturity;
2. Any technologically mature civilization is extremely unlikely to run a significant number of computer simulations of their evolutionary history (or variations thereof);
3. We are almost certainly living in a computer simulation. (Bostrom, 2003, p. 14).

Extrapolating from tendencies and preferences that have been defining how humans currently develop and use digital media, Bostrom imagines a civilization that reached the technical capability "to convert planets and other astronomical resources into enormously powerful computers" (ibid., p. 3). In that hypothetical scenario, and should that civilization maintain sufficient interest in developing and running what Bostrom calls 'ancestor-simulations'⁴, then – he argues – it is almost a statistical

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/a-theodicy-for-artificial-universes/269433

Related Content

Increasing Transparency When Assessing the Impact of Technology on Human Well-Being: A Capability Approach Perspective

Julien Forbat and Ingrid Vargas (2021). *International Journal of Technoethics* (pp. 32-53).

www.irma-international.org/article/increasing-transparency-when-assessing-the-impact-of-technology-on-human-well-being/269434

An Access Control Framework for WS-BPEL Processes

Federica Paci, Elisa Bertino and Jason Crampton (2013). *Digital Rights Management: Concepts, Methodologies, Tools, and Applications* (pp. 310-334).

www.irma-international.org/chapter/access-control-framework-bpel-processes/70982

Societal Implications of Current and Emerging "Smart" Technologies

Octavian Mihai Machidon (2015). *International Journal of Technoethics* (pp. 60-70).

www.irma-international.org/article/societal-implications-of-current-and-emerging-smart-technologies/124868

Recent Copyright Protection Schemes: Implications for Sharing Digital Information

Herman T. Tavani (2005). *Intellectual Property Rights in a Networked World: Theory and Practice* (pp. 182-204).

www.irma-international.org/chapter/recent-copyright-protection-schemes/24119

De-Marginalizing Technophilosophy and Ethical Inquiry for an Evolving Technological Society

Rocci Luppigini (2018). *International Journal of Technoethics* (pp. 1-14).

www.irma-international.org/article/de-marginalizing-technophilosophy-and-ethical-inquiry-for-an-evolving-technological-society/208945