

This paper, written by Nick Bostrom, proposes that at least one of the following statements must be true:

- The human species is very likely to go extinct before reaching a “posthuman” stage.
- Any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history.
- We are almost certainly living in a computer simulation.

If the case that we are currently living in a simulation is true, then the belief that there is a significant chance that we will one day become posthumans who run ancestor-simulations is also true. Additionally, the paper also outlines various other effects of this outcome.

To begin, the paper puts forth a series of events and logical statements in order to further understand the premise of his thesis. If we suppose that the propositions—which are made by science fiction, technologists, and futurologists—revolving around the likelihood of having access to mass amounts of computing power in our future is true, then the possibility of those future generations running precise simulations of either their ancestors or people similar to their ancestors is also true. Further, if we assume the ability to run mass amounts of simulations is probable and that the people in these simulations are conscious, then it is possible to believe that all of those conscious people’s minds, such as our own, belong to the descendants of an original race and not their own. If we suppose we do belong to a computer simulation, we are qualified to believe that we will have future generations who obtain the capabilities of conducting such simulations as discussed above. Additionally, the paper notes that from this thesis one may be led to a series of interesting and intriguing questions.

Next, the paper brings up a philosophical concept called substrate-independence, which is related to consciousness, and addresses the thesis of the acceptability of consciousness being implemented by a computer; this paper assumes such a thesis is true. Further, this paper does not assume that in order to create a mind on a computer it is not necessary to have it act like a human in all cases, rather it assumes a weaker proposition.

This section of the paper starts by discussing how, because of the constant and rapid technological process of our world, it is very possible that if this progress continues, we could have hardware and software powerful enough to replicate the human consciousness in a computer. The paper goes on to say that although we know that this is true, it is impossible to come up with an accurate timeline for when this will happen, because of a number of factors including ones that may be unforeseen such as undiscovered physical phenomena. After this the paper discusses the computation power of currently known machines, and contrasts this with the amount of computing power that the human brain uses. The paper lists several factors that affect the computing power of the human brain that would be necessary to accommodate for, in a simulation of the brain such as a central nervous system, memory, and environmental responses. The paper then goes on to discuss the

possibility of an environmental simulation, and it lists factors that are necessary for that. The paper states that it would not be necessary to include every microscopic detail exactly but it would be necessary to have a simulation where when you look through a microscope it is unsuspecting. It also states that the main computational cost in creating a simulation that is indistinguishable from reality to a human is in creating accurate simulated brains. In a “postman world” creating this type of simulation could be entirely possible.

The paper then goes on to ask the question, “if we know that this will be possible in the future, how do we know that we are not currently in an ancestor simulation?” This question is answered in the form of the three statements previously discussed, where at least one of them must be true:

- The fraction of all human-level technological civilizations that survive to reach a posthuman stage is essentially zero.
- The fraction of posthuman civilizations that are interested in running ancestor-simulations (or that contain at least some individuals who are interested in that and have sufficient resources to run a significant number of such simulations) is essentially zero.
- The fraction of all people with our kind of experiences that are living in a simulation is very close to one.

The fifth section of this article discusses the bland indifference principle. The idea is given a lack of information on a subject matter, if we have a number of possibilities to choose from, how can we determine which one is likely to occur? The bland indifference principle says to assign equal probabilities to each situation. This way we don’t have to decide one over the other. Bostrom shows two examples of the indifference principle’s effectiveness where there is lack of information. When dealing with the question as to whether we live in a simulator, the first case says to treat all the minds in question as our own. This case is nice because you know what your mind is like, so comparing to other minds through yours is quite convenient. The second case takes into account the uniqueness of other people’s minds. While the latter case may not sound as convenient as the first case, Bostrom argues that this is okay, and that the simulation argument still works. To argue that the second case works, he uses biology. More specifically, he says that if a  $x\%$  of a population has the same genetic sequence  $S$  in their DNA and that there are no manifestations of that genetic sequence, then it is logically sound to assign a credence of  $x\%$  to the hypothesis— the hypothesis being you have the genetic sequence  $S$ . Bostrom goes on to note that the indifference principle suggests indifference between hypotheses considering you are the observer and that you have no information pertaining to which observer you may be. He finishes this section by arguing that the bland indifference principle is a similar assumption used before the Doomsday argument.

The sixth section describes interpretation. The kind of interpretation Bostrom is discussing is the kind in which unpacks several circumstances that may occur if one of the propositions were true. He notes that the first proposition by itself does not necessarily allude to the fact that humans will go extinct soon. Of course, there are some natu-

ral interpretations to consider when discussing these propositions. One such interpretation is that we as humans may go extinct because of some dangerous, yet powerful technology. Furthermore, he shares another aspect in the conclusion of the simulation argument; that is, the fraction of posthuman folks who want to create an ancestor-simulation is quite small. The only way this can be considered true is if there is some strong convergence. Bostrom brings to mind what types of ideas may not lead posthuman civilizations to create an ancestor-simulation. One type of idea is post-human's ethical views. Namely, having an ancestor-simulation may be immoral. Another convergence point could be that posthumans lose interest in running an ancestor-simulation. As for the third possibility (or proposition), Bostrom suggests if we are indeed in a simulator, then the physics of our universe could be different from the physics of the universe the computer is run on. He compares this kind of thinking to that of virtual machines, where JavaScript applets run on the virtual machine. Not only does he give comparisons to that of computer science concepts, but also to that of religious ideas. For instance, he shares how posthumans who are running the simulation are gods—namely, they have superior intelligence and can be considered omnipotent. Bostrom comments if one were to study this sort of idea, particularly the hierarchical levels of reality, then one would be led to the study of naturalistic theogony. He wraps up the discussion in the section by denoting that if we do live in a simulation, the implications for humans are not as bad as they may seem.

The seventh section concludes his remarks on the tripartite conclusions; namely, that due to empirical fact, at least one of the propositions mentioned are true. If one is true, then there is a high chance we as humans will go extinct before reaching posthumanity. If two is true, then strong convergence must interfere with the plans of the wealthy or entitled to host an ancestor-simulation. And as for three, if it were true, then we most likely live in a simulator. Bostrom declares if we are in a simulator, then it is safe to assume that our descendants will not run an ancestor simulation.