

# The Simulation Argument: Why the Probability That You Are Living in a Matrix is Quite High

by Nick Bostrom (*Times Higher Education Supplement*, May 16, 2003)

---

This is a popular piece summarizing Bostrom's academic article:

Bostrom, Nick (2003). "Are We Living in a Computer Simulation?" *Philosophical Quarterly* 53(211).

---

The Matrix got many otherwise not-so-philosophical minds ruminating on the nature of reality. But the scenario depicted in the movie is ridiculous: human brains being kept in tanks by intelligent machines just to produce power.

There is, however, a related scenario that is more plausible and a serious line of reasoning that leads from the possibility of this scenario to a striking conclusion about the world we live in. I call this the simulation argument. Perhaps its most startling lesson is that there is a significant probability that you are living in computer simulation. I mean this literally: if the simulation hypothesis is true, you exist in a virtual reality simulated in a computer built by some advanced civilisation. Your brain, too, is merely a part of that simulation. What grounds could we have for taking this hypothesis seriously? Before getting to the gist of the simulation argument, let us consider some of its preliminaries. One of these is the assumption of "substrate independence". This is the idea that conscious minds could in principle be implemented not only on carbon-based biological neurons (such as those inside your head) but also on some other computational substrate such as silicon-based processors.

Of course, the computers we have today are not powerful enough to run the computational processes that take place in your brain. Even if they were, we wouldn't know how to program them to do it. But ultimately, what allows you to have conscious experiences is not the fact that your brain is made of squishy, biological matter but rather that it implements a certain computational architecture. This assumption is quite widely (although not universally) accepted among cognitive scientists and philosophers of mind. For the purposes of this article, we shall take it for granted.

Given substrate independence, it is in principle possible to implement a human mind on a sufficiently fast computer. Doing so would require very powerful hardware that we do not yet have. It would also require advanced programming abilities, or sophisticated ways of making a very detailed scan of a human brain

that could then be uploaded to the computer. Although we will not be able to do this in the near future, the difficulty appears to be merely technical. There is no known physical law or material constraint that would prevent a sufficiently technologically advanced civilisation from implementing human minds in computers.

Our second preliminary is that we can estimate, at least roughly, how much computing power it would take to implement a human mind along with a virtual reality that would seem completely realistic for it to interact with. Furthermore, we can establish lower bounds on how powerful the computers of an advanced civilisation could be.

Technological futurists have already produced designs for physically possible computers that could be built using advanced molecular manufacturing technology. The upshot of such an analysis is that a technologically mature civilisation that has developed at least those technologies that we already know are physically possible would be able to build computers powerful enough to run an astronomical number of human-like minds, even if only a tiny fraction of their resources was used for that purpose.

If you are such a simulated mind, there might be no direct observational way for you to tell; the virtual reality that you would be living in would look and feel perfectly real. But all that this shows, so far, is that you could never be completely sure that you are not living in a simulation. This result is only moderately interesting. You could still regard the simulation hypothesis as too improbable to be taken seriously.

Now we get to the core of the simulation argument. This does not purport to demonstrate that you are in a simulation. Instead, it shows that we should accept as true at least one of the following three propositions:

- (1) The chances that a species at our current level of development can avoid going extinct before becoming technologically mature is negligibly small.
- (2) Almost no technologically mature civilisations are interested in running computer simulations of minds like ours.
- (3) You are almost certainly in a simulation.

Each of these three propositions may be *prima facie* implausible; yet, if the simulation argument is correct, at least one is true (it does not tell us which).

While the full simulation argument employs some probability theory and formalism, the gist of it can be understood in intuitive terms. Suppose that proposition (1) is false. Then a significant fraction of all species at our level of development eventually becomes technologically mature. Suppose, further, that (2) is false, too. Then some significant fraction of these species that have become technologically mature will use some portion of their computational resources to run computer simulations of minds like ours. But, as we saw earlier, the number of simulated minds that any such technologically mature civilisation could run is astronomically huge.

Therefore, if both (1) and (2) are false, there will be an astronomically huge number of simulated minds like ours. If we work out the numbers, we find that there would be vastly many more such simulated minds than there would be non-simulated minds running on organic brains. In other words, almost all minds like yours, having the kinds of experiences that you have, would be simulated rather than biological. Therefore, by a very weak principle of indifference, you would have to think that you are probably one of these simulated minds rather than one of the exceptional ones that are running on biological neurons.

So if you think that (1) and (2) are both false, you should accept (3). It is not coherent to reject all three propositions. In reality, we do not have much specific information to tell us which of the three propositions might be true. In this situation, it might be reasonable to distribute our credence roughly evenly between the three possibilities, giving each of them a substantial probability.

Let us consider the options in a little more detail. Possibility (1) is relatively straightforward. For example, maybe there is some highly dangerous technology that every sufficiently advanced civilization develops, and which then destroys them. Let us hope that this is not the case.

Possibility (2) requires that there is a strong convergence among all sufficiently advanced civilisations: almost none of them is interested in running computer simulations of minds like ours, and almost none of them contains any relatively wealthy individuals who are interested in doing that and are free to act on their desires. One can imagine various reasons that may lead some civilisations to forgo running simulations, but for (2) to obtain, virtually all civilisations would have to do that. If this were true, it would constitute an interesting constraint on the future evolution of advanced intelligent life.

The third possibility is the philosophically most intriguing. If (3) is correct, you are almost certainly now living in computer simulation that was created by some advanced civilisation. What kind of empirical implications would this have? How should it change the way you live your life?

Your first reaction might think that if (3) is true, then all bets are off, and that one would go crazy if one seriously thought that one was living in a simulation.

To reason thus would be an error. Even if we were in a simulation, the best way to predict what would happen next in our simulation is still the ordinary methods – extrapolation of past trends, scientific modelling, common sense and so on. To a first approximation, if you thought you were in a simulation, you should get on with your life in much the same way as if you were convinced that you are living a non-simulated life at the bottom level of reality.

The simulation hypothesis, however, may have some subtle effects on rational everyday behaviour. To the extent that you think that you understand the motives of the simulators, you can use that understanding to predict what will happen in the simulated world they created. If you think that there is a chance that the simulator of this world happens to be, say, a true-to-faith descendant of some contemporary Christian fundamentalist, you might conjecture that he or she has set up the simulation in such a way that the simulated beings will be rewarded or punished according to Christian moral criteria. An afterlife would, of course, be a real possibility for a simulated creature (who could either be continued in a different simulation after her death or even be “uploaded” into the simulator’s universe and perhaps be provided with an artificial body there). Your fate in that afterlife could be made to depend on how you behaved in your present simulated incarnation. Other possible reasons for running simulations include the artistic, scientific or recreational. In the absence of grounds for expecting one kind of simulation rather than another, however, we have to fall back on the ordinary empirical methods for getting about in the world.

If we are in a simulation, is it possible that we could know that for certain? If the simulators don’t want us to find out, we probably never will. But if they choose to reveal themselves, they could certainly do so. Maybe a window informing you of the fact would pop up in front of you, or maybe they would “upload” you into their world. Another event that would let us conclude with a very high degree of confidence that we are in a simulation is if we ever reach the point where we are about to switch on our own simulations. If we start running simulations, that would be very strong evidence against (1) and (2). That would leave us with only (3).