

## READING GUIDE FOR “THE SIMULATION ARGUMENT: WHY THE PROBABILITY THAT YOU ARE LIVING IN A MATRIX IS QUITE HIGH”

The simulation that is the topic of this article is sort of like a computer game, except that the individuals in it are conscious—that is, they each have an inner life—and the game is completely life-like. It’s indistinguishable from what we experience every day. Unlike a computer game, however, there are no “players” who control each specific movement of the people in the simulation. Someone (or a group) would have created the simulation, and then they just let it run—although if they wanted to interfere with the simulation, presumably, they could at any time.

There’s more on this below, but if we—and our whole world—exist in a computer simulation, then the creators of the simulation are not necessarily human beings, although they are members of an intelligent species. If they created us to look just like them, then they would be human or something similar, but they could also be a completely different kind of species.

Throughout this article, Bostrom (the author) refers to “simulated minds” rather than simulated people. That’s mainly for two reasons. First, for a computer simulation, simulating a conscious mind is the real task—simulating everything else in our world would be relatively easy. Second, as Descartes said, what you are is your mind, and so, when referring to people (real or simulated), Bostrom talks about minds. But the idea is that you and everyone else—in or out of the simulation—has a body and exist in a world (just like you think you do).



On the first page—in particular, for paragraphs two, four, and five—try to identify the main point of each paragraph. Notice that paragraph five begins with “Our second preliminary.” The first preliminary is first mentioned in paragraph two. Make sure you’re clear on those two issues (those two preliminaries) that Bostrom discusses on the first page.

There’s a lot packed into the idea of “substrate independence.” Bostrom explains what it means, and you’ve had some exposure to this topic from reading the chapter on mind. In brief, “substrate independence” is the idea that the *substrate*

(or material) that a conscious mind is made of doesn't matter. Even if this seems like a big (and perhaps unusual) idea, in the chapter on mind, we found that dualism doesn't work as a theory of the mind. That just leaves monism—the idea that everything, including the mind, is made of physical material—and that's all that's needed to claim that substrate independence is possible. It might take thousands of years to figure out how to build the computers and write the programs that can run the “processes that take place in your brain.” But however long it takes doesn't matter for Bostrom's argument.

At the end of paragraph five (on the first page), Bostrom says,

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.

Regarding **a technologically mature civilization**, Bostrom is talking about a civilization that is far beyond ours in terms of the technology that they've developed. It could be a civilization that is thousands of years (or tens of thousands years) beyond us. But if they began much earlier, then they would have reached a technologically advanced stage sometime in the past.

This is also assuming that there is other intelligent life in the universe—not in our solar system and not necessarily in our galaxy—but in the universe. The crucial step for making the claim that there is other intelligent life out there is realizing that the universe is really, really big. Galaxies are huge. There are tens of billions of solar systems in our galaxy, and there are over 100 billion galaxies in the universe. Thus, the possibility of other intelligent life somewhere is quite high. Even if the probability that any particular planet in the universe has intelligent life on it is miniscule, there are so many planets, that it's almost 100 percent certain that there's some other intelligent life in the universe.

Here's the background for an **astronomical number of human-like minds**. If a technologically advanced civilization, can run a computer simulation of a world with conscious, intelligent life, then presumably it would create many, many

people. If one world like ours could be created in a computer simulation, then it seems just as likely that multiple worlds like ours would be created, maybe even lots. (There are, right now, 7 billion people on earth. If running a simulation of a world like ours was no more complex for a technologically advanced civilization than, say, running an email or data server is for us, then they could be running thousands—7 billion  $\times$  5,000 different simulations = 30 trillion people. If it's as simple as running a video game, then it could be a lot more.) Plus, if there are multiple technologically advanced civilizations in different parts of the universe, then it would be 30 trillion (or whatever) times the number of technologically advanced civilizations. Anyway, point being—and this is going to be important on page two of the article—if there are any computer simulated people, then it's likely that there will be “an astronomical number of human-like minds” that are all just in one computer simulation or another.



On page two, (1), (2), and (3) are central to Bostrom's argument, which basically comes down to the reasons that we have (or don't have) for thinking that each one is true or false. Statement (1) is worded as a probability, but that's not especially important. Basically, (1) is “species like us go extinct before reaching a technologically advanced stage.” The species in question are other lifeforms anywhere in the universe, and so the idea is—if (1) is true—we, and all other species that are at our stage of technological development, go extinct before we can run these computer simulations.

In the first half of this page, Bostrom discusses the reasons why (1) and (2) might be false, and at the bottom of the page, he discusses reasons why they might be true.



If (1) is false, then species like us do *not* go extinct. If (1) and (2) are false, then species make it to this technologically advanced stage and they have no problem running these simulations containing conscious people. The next step is in the middle of page two. Bostrom states,

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.

We went through the math above. An advanced civilization could create trillions of simulated people (or even more). There have to be some non-simulated people out there (i.e., real people), however, and so we could be real. But, non-simulated people are going to be a tiny fraction of all the people with conscious minds (those in simulations and those not in simulations). Hence, for any person, if you're just going with the odds, then you probably exist in a computer simulation—or rather, all of us in this world are probably in a computer simulation. That is, if (1) and (2) are false.

The *principle of indifference* is, basically, as I said right above, “just going with the odds.” More technically it's this: if there are multiple alternatives and there are no reasons that make any of them more or less likely, then an equal probability should be assigned to each alternative.

In the situation at hand, if there are, in the universe, say, 99 simulated conscious minds for every non-simulated one, then the probability of being a non-simulated person is one percent. But since the simulation is just a computer program run by a technologically advanced civilization, the ratio could 999 to 1 (one-tenth of one percent) or 9,999 to 1 (one one-hundredth of one percent). So, “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” (paragraph 9).



Re-read the fourth paragraph on page two. That's the conclusion of the simulation argument.

On the one hand, if both (1) and (2) are false, then it seems that we have to accept (3), we're living in a computer simulation. Moreover, for other reasons, we might hope that (1) is false—namely, we don't want our species to go extinct.

Similarly, just considering (2) on its own, we might think it's false because, given the opportunity, we, or at least someone in the U.S., would probably run such a simulation if it were currently possible.

On the other hand, Bostrom says, "We do not have much specific information to tell us which of the three propositions might be true." Thus, we can just assign a  $1/3$  chance of being true to each of (1), (2), and (3).



In the first two paragraphs on page three, Bostrom points out that, even if he has established that there is a real chance that we are in a simulation, that doesn't really change anything about our lives. Notice that this is, in a way, similar to the perspective we take on the earth orbiting the sun as well as the possibility that determinism is true. But in the second-to-last paragraph, Bostrom does suggest that, even if our day-to-day lives are unchanged, the motives of the simulators might have implications for us. Then, in the last paragraph, he describes two situations that would give us quite a bit of evidence that we are living in a simulation.