This essay will be examining the Simulation Argument put forward by Nick Bostrom in his article, *Are we Living in a Computer Simulation*. Each premise will be analysed individually, then the overall validity of the argument and its implications.

Firstly, the argument rests on the assumption that human minds can be simulated. There are a few components to this: It is possible with enough computing power to simulate a human mind in its entirety, a simulated mind will be conscious, and an advanced civilisation could acquire this level of computing power. If there were good reasons to assume that the human mind was capable of feats of computation that a computer simply could not replicate, then this would debunk the simulation argument. This is because the parts of the mind with this property could not be replicated by a simulation of a brain and, because we have this property, it would be impossible that we are simulated. It has been argued that the brain is a kind of hyper-computer that can compute things a regular computer cannot, such as whether or not a program will ever stop running (Lucas, 1961), but such arguments have been convincing disproved by many. (Campbell, 2016).

Whether such a simulated mind is conscious or not rests on the truth of Functionalism, as it needs to be possible for a conscious mind to arise out of pure computation. Functionalism is a very widely accepted belief (Bostrom, 2003) and has many reasons to believe it, while having little in terms of convincing counter-arguments to it. Both the Chinese room and Chinese nation argument have been used as an attempt to debunk functionalism, while both are rather absurd notions, neither are convincingly implausible. Complicated behaviour arising from simple systems is extremely common in nature, such as in ant or bee colonies, where the whole is far more complex than the sum of its parts. It should also be acknowledged that a very common belief within, in particular, religious communities and other similar sects of society is that there is an ethereal component to human nature and consciousness, the soul. If such a thing were to exist it is hard to imagine it could be replicated by a computer, and so the argument would no longer be plausible. However, no physical evidence currently points to the existence of a soul and therefore is not a good reason against the soundness of the argument. Looking purely at the chance of the average person accepting or rejecting certain premises, then this is the weakest one in the argument. Even so this a widely accepted notion that is difficult to argue against.

The third assumption is that it is theoretically possible for future civilisations to acquire enough computing power to run these simulations. Bostrom estimates that between 1033 and 1036 operations would be needed to simulate 100 billion humans for 50 years, this number isn’t including the computation needed for simulated the environment, but it is reasonable that this cost would be negligible compared to the brains of humans. This is assuming that the parts of the world that are not being observed can be heavily compressed and still function as if they weren’t once they are observed. It’s hard to imagine this would be very difficult to do at all for a civilisation as advanced as we’re imagining. Being extremely conservative and assuming simulating the environment takes 1000 times as many operations as the brains would, the absolute most amount of operations a simulation would take is 1039 operations. This shouldn’t even be needed as there’s not much reason to think the cost of simulating the environment is at all significant next to the cost of simulating humanities consciousness. A planetary mass computer using only our current technology is roughly estimated to be able to carry out 1042 operations per second, meaning even with the massively overestimated cost of 1039 operations per simulation, such a computer would be able to carry out 1000 simulations in just a single second. In reality, this would likely be millions of times easier with the realistic ease of simulating the environment and how superior the future technology would be. With technology like quantum computing (something that we are even able to make use of now, although in a severely limited capacity) would shrink the theoretical size of a computer with the same power astronomically. Even so, building planetary size computers is theoretically trivial for a post-human civilisation and so such a civilisation would probably build millions or even billions of such computers. And so even if a minute fraction of their computing power is used for these simulations, there could still be trillions of simulations being run. Even with massive margins of errors, the assumption that a posthuman civilisation would have enough computing power to run these simulations is extremely likely and difficult to dispute.

The argument also assumes a weak indifference principle, effectively stating that if that we (or other civilisations and simulations) have no indication of whether or not we are in a simulation, and there is an x% chance of any particular civilisation being in a simulation, then we have an x% chance of being in a simulation. This appears to be a logical truism, there isn’t much that could be said to counter such a statement. However, for this to work for the simulation argument there would need to be no indication of whether or not we are in a simulation. Perhaps there is reason found within the nature of reality that points to it being a simulation, many unanswered questions that seem almost impossible could be blamed on it simply being how the simulation was designed, such as why the speed of light is 3x108 ms-1. Ultimately though such uncertainties only increase our chance of being in a simulation, not diminish it. There doesn’t seem to be any particular evidence to doubt we are in a simulation. So, the indifference principle the argument is relying on appears to be completely valid.

Bostrom argues that following the aforementioned assumptions, a formula can be deduced to find the probability that we are in a simulation:

Where fsim is our chance of being in a simulation, fp is the fraction of civilisations that survive to a posthuman stage, f1 is the fraction of posthuman civilisations that would be interested in running simulations, and N1 is the number of simulations such a civilisation would run.

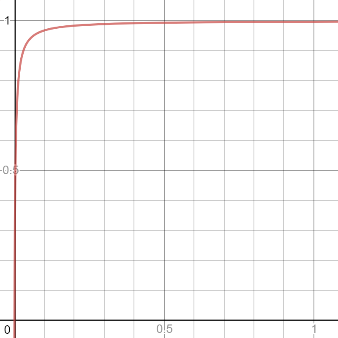
Bostrom concludes that either fsim is effectively one, fp is effectively zero, or f1 is effectively zero.

This is because Bostrom believes that N1 would be astronomically large. Using the low estimate of a planetary sized computer and the (unfairly) high estimate of the cost of a simulation, it would take such a computer diverting all of its resources one second to simulate 1000 realities, if this one computer ran for a year it would create 3 x 1010 or 30 billion realities. Given that this computer would likely be a minute fraction of the total computing power this society has access to and that the societies total life span in which they run these simulations would far, far exceed that of just a year, then 30 billion is an extremely conservative estimate for N1, and so N1 is astronomically large.

There are lots of reasons to assume that the fraction of civilisations that survive to a posthuman stage, fp, is very small. As our society grows more advanced and masters more technologies, the easier it is becoming to wipe ourselves out. This could be done now quite spectacularly with nuclear weapons if the people with access to them were so inclined, and we face similar threats with the development of GNR technologies. A civilisation might also choose to stunt their development permanently before reaching a post-human stage, due to the fear of being shut down if they were in a simulation, as simulating a post-human simulation would take an enormous amount more computing power than we considered earlier, so much so that the creators of the simulation might choose to terminate the simulation because of the amount of computational power it is taking up. It certainly isn’t good news if this is close to zero as there is no reason to think that we are a special case when it comes to overcoming these threats.

There could be many reasons a post-human civilisation wouldn’t run ancestor simulations, perhaps there is some ethical reason as to why it is immoral to run these simulations and so it is outlawed in an effectively enforced way. Or these post-humans could simply just not care to run such simulations, this would mean these post-humans would be significantly different to us, as there are many reasons these would be run today if they could, in fields like science, history, entertainment, and much more. But prescribing these same desires to beings that are incredibly more intellectually and technologically superior to us is a fool’s errand. So, f1 could be close to zero.

Because of the enormous size of N1, as long as the chance of fp or f1 isn’t effectively zero, then there is a very high probability that we are currently living in a simulation. Using the low estimate of N1 of 3 x 1010 and prescribing a low (but not close to 0) chance of only 0.1% (0.001) to both fp and f1, the chance of us currently living in a simulation is 99.996% which is only a chance of 1 in 100000 that we are not living in a simulation.

This is what a graph of the above equation looks like, with the vertical axis being the probability of being in a simulation (fsim), and the horizontal axis being the probability of extinction before the post-human stage (fp). Using the chance of a civilisation running simulations being only 1 in a billion. Even with such a negligible number, at just a 10% chance of reaching the post-human stage, fsim still has a 97% likelihood, with fp at 1%, fsim is 75%, and when fp is a terrifyingly small number of only 0.1%, there’s still a 23% chance (almost a quarter) that we are indeed living in a simulation. (The graph and other values would look and function the same if the values of fp and f1 were swapped.)

The existence of the argument can mean several things, all of which challenge the beliefs of most people. The rejection of the argument as a whole is to reject the truth of at least 1 assumption, maybe the brain is a hyper-computer, functionalism is wrong, there could be an unforeseen blockage in the possible power of predicted computers that severely limits their theoretical limits, our understanding of statistics and math might even be fundamentally wrong. Either way, given the many good reasons to believe in the assumptions, the rejection of any would be at best extremely surprising to most people, and at worst totally reality shattering. To accept any of the 3 conclusions is hardly more comfortable, the best-case scenario is that a post-human civilisation wouldn’t want to run simulations, this would imply that humanity would reach a point that they are completely unrecognisable to us, it would be like talking to aliens. Society would be shifted so dramatically that even pondering it would be like an ant trying to understand a nuclear reactor. Alternatively, humanity is certain to be wiped out before technologically maturing, an existentially terrifying idea. Or finally, all of reality is data on a computer perpetually watched and analysed by a future civilisation.

Overall there is little reason to reject any of the premises and the conclusions follow from the premises. The argument as a whole is very sound.

# Bibliography

Bostrom, N. (2003, April). Are We Living in a Computer Simulation? *The Philosophical Quarterly, 53*(211), 243-255.

Campbell, D. (2016). Why We Shouldn’t Reason Classically, and the Implications for Artificial Intelligence. *Computing and Philosophy, 375*, 151-165.

Lucas, J. R. (1961). Minds, Machines and Gödel. *Philosophy, 36*(137), 112-127. doi:10.1017/S0031819100057983