**The simplicity assumption and some implications of the simulation argument for the future of humanity**

Lorenzo Pieri

**Abstract**

According to the most common interpretation of the simulation argument, we are very likely to live in an ancestor simulation. It is interesting to ask if a priori some families of simulations are more likely than others inside the space of all simulations. We argue that a natural probability estimate is given by computational complexity: easier simulations are more likely to be runned. Remarkably this allows us to extract experimental predictions from the fact that we live in a simulation. For instance we show that it is very likely that humanity will not achieve interstellar travel, that humanity is not going to meet other intelligent species in the universe and that older universes are less likely than younger ones. On the opposite side, experimental falsification of any of these predictions would constitute evidence against our reality being a simulation.

# Simulation Argument

The simulation argument [1] is really a trilemma, in which our universe being a simulation appears to be the most likely option. In a nutshell, the argument says that given our ever increasing ability to run powerful simulations and no apparent physical limit, we are led to conclude that we will be able to build civilization-size simulations. If it's possible, then likely someone else already did it, and the likelihood of us being the original simulators is slim. In the following paper we will assume that we are indeed in a simulation and explore some of the observational consequences of being inside a simulation.

# The simplicity assumption

How can we infer anything about our reality, just from the fact that our reality is a simulation? There is no a priori way to do this exactly, but we can make probabilistic arguments, and in fact the simulation argument itself is a probabilistic argument. In particular, in the space of all possible simulations we are very likely to find ourselves among the most likely classes of simulations[1]. From our everyday experience we know that those simulations should at least be powerful enough to be able to simulate our solar system and our (apparently?) conscious experience in great detail.

In a typical multiplayer video game setting, environments close to the players are rendered with high precision, while distant sections of the game universe are only approximated. Given the current status of humanity space colonization, we can argue that simulating in detail only the entire solar system would be compatible with our observations.

So now the question is, what is the most likely class of simulations compatible with the observations of a civilization?

An answer is given by the following assumption:

*The simplicity assumption (SA): If we randomly select a simulation in the space of all possible simulations of a civilization that have ever been runned, the likelihood of picking a given simulation is inversely correlated to the computational complexity of the simulation.*

In a nutshell, simpler simulations are more likely to be runned.

Suppose that simulation A takes 1000 times more elementary computations than simulation B, to simulate the same civilization. How much more likely is B with respect to A? Here we give some heuristic arguments to show that B is exponentially more likely:

It doesn't matter how large, the simulators must have finite computational power at their disposal. Here by simulators we mean all the simulators existing at any given time that are running simulations of our civilisation, including simulated simulators. Suppose that at $t_{start}$ the simulators obtain for the first time enough computational power to run our civilization in the simplest possible way. Without loss of generality we call the associated computational cost per unit of time $c_1 = 1$. The simulators maintain this simulation power for a period $t_1$, after which they obtain a superior computational power per unit of time (which for simplicity we assume being a multiple of $c_1$, see the appendix for details) $c_2 = 2\,c_1$, which they will maintain for a period $t_2$. At this point, in every unit of time, they can have the option of allocating computational resources to a single simulation of computational cost $c_2$ or two simulations of computational cost $c_1$. A priori at any given time we have no reason to believe that allocating computationally resources in one way should be preferred, as we are considering the space of all possible simulations that have ever been runned, which may have been run by a large number of unrelated simulators, perhaps some of them even in

[1] Here we implicitly assume the SIA (Self-Indication Assumption): All other things equal, an observer should reason as if they are randomly selected from the set of all *possible* observers. In this case, the observer is an entire simulated civilization. [2]

different nested levels of simulations. So we conclude that all possible ways in which the simulations can be run are sampled uniformly over many units of time.

This imply that at large $n$ for every instance in which a simulation of cost $c_n$ is run there are also $n_1$ instances of $c_1$ simulations, where

$$n_1 \approx \frac{e^{\pi\sqrt{\frac{2}{3}(n-1)}}}{(n-1)}$$

Also, simulations costing $c_1$ were also runned for all the $t_k$ with $k < n$.

In a nutshell, simple simulations are much more likely since for a long time they are the only possible simulations that can be runned, and later they can be runned a huge amount of times with respect to the limited number of times for a more costly simulation.

On top of this we notice that:

- A possible use of civilization simulations (and one that we would anthropocentrically consider likely) is scientific research. To achieve high statistical significance a simulation must be run a large number of times. A rational simulating scientist would settle on the simpler simulation that is complex enough to feature all the elements of interest and then run that simulation over and over.
- Simple simulations are the only simulations that can be runned inside nested simulations, due to limits in computational power.

An example partially illustrating the above points are the first classic arcade video games. Not only have these games been played billions of times, they have also been featured inside larger video games and used as test benchmarks for training reinforcement learning and artificial intelligence algorithms.

To obtain stronger conclusions and since we are interested in a lower bound, instead of assuming exponentially scaling, in the remainder of the paper we will limit to demand that the scaling implied by the simplicity assumption is at least linear. We call this SA with linear scaling the *Weak Simplicity Assumption*, and also define the *Strong Simplicity Assumption* as the SA with exponential scaling. So, if we randomly select a simulation, B is about 1000 times more likely than A to be picked using the Weak SA, while is about $10^{32}$ more likely using the Strong SA.

In this paper we will further assume that, simulated or not, our reality does not arise from a Boltzmann Brain, a Brain-in-a-Vat or other kinds of solipsistic universes. Notice that if we don't explicitly assume it, the SA is actually implying that we are overwhelmingly likely to be one of such brains or “solo players”, as it is much easier to simulate the inputs to the brain than the full blown reality.

So far when speaking about complexity we referred to time computational complexity. A more refined argument can be made including other kinds of complexities, for instance space computational complexity and the Kolmogorov complexity of the program running the universe. The latter addition looks particularly suited in the case in which the simulators are not intelligent, for instance in the case in which the computer running our simulation simply emerged as random fluctuation. In this “simulation without simulators” scenario, longer programs are much less likely than shorter ones since they need to emerge randomly from the space of all computer programs.

# Modelling the simulation

How is the simulation implemented in code? We cannot directly answer this question, but we only need to know how the computational complexity of the simulation scales to obtain the relative likelihood of two simulations, factoring out our ignorance of the details.

Suppose that the simulation is composed of atomic entities, or atoms. These may be actual atoms, more elementary quantum fields, strings or branes, it doesn’t matter. What we know is that the computational complexity of the simulation will be positively correlated with the number of atoms. What is the exact scaling relation? We don’t know, but we can put a lower bound by drastically simplifying all the quantum and gravity interactions into a simple N-body simulation problem and also assuming an extreme level of algorithmic efficiency of the simulators, namely the simulation can be at best O(N) in the number of atoms. For instance the fastest N-body algorithms used in astrophysics, such as the Fast Multipole Method, can be O(N) for a given precision.

The computational complexity of simulations is therefore at best linearly proportional to the time the simulation is runned, multiplied by the number of atoms. In the following we will consider actual atoms (hydrogen, etc.), as their number is positively correlated with the fundamental atomic entities.

We have finally set the stage and we are ready to draw some observables consequences of living in a simulation.

# Interstellar Travel

A typical strategy to limit the computational cost of open ended virtual worlds is to accurately simulate local physics and strongly approximate the physics at the horizon. An efficient simulation of our solar system would simulate no more than 1-2 light years centered around the sun, therefore approximating every other star that we can see in the night sky[2].

Stars are the dominant sources of atoms density in our neighbourhood, as the density of interstellar space is only about 0.05 atoms/cm^3 in the Local Bubble and 0.5 atoms/cm^3 in the interstellar medium of the Milky Way. Our solar system mass for instance is 99.85% concentrated in the sun.

There are hundreds of stars in the 25Lys around the sun, with a total mass in the order of $10^2$ solar masses [3]. Even disregarding dark matter and interstellar gas, this means that a simulation of our universe with full rendering of the closest 25Lys is no less than 100 times more computational intensive with respect to a simulation in which we are confined to our solar system, and therefore 100 times less likely according to the Weak SA and having very efficient algorithms simulating the universe.

If we move further, the milky way has more than 100 Billion stars and 100 millions of black holes, with a visible mass $$10^{12}$$ the one of our solar system, making a simulation in which humanity is able to perform interstellar travel extremely unlikely, regardless of which SA we use.

In summary, the simulation argument combined with the simplicity assumption predicts the absence of significant interstellar travel for our civilisation (or the invention of von Neumann probes and other means of exploring large portions of space).

# Extraterrestrial Intelligence

As a corollary, even with the mildest scaling assumptions there is a probability no larger than 1% of contacting other intelligent species out there. Our best bet is to find them in very nearby stars.

Fermi's paradox is therefore solved: we don't see aliens, since we live in an efficient simulation in which the majority of habitable planets are too far away.

Here we are assuming that there are no hidden aliens in some cave on mars, and that aliens are not "dummy characters", but they have similar experiences to us. Basically in the simulation no intelligent observers are preferred. Simulating far away aliens requires

---

[2] The simulation's approximation to far away physics should emerge as inconsistencies in cosmological measurements. It is tempting to speculate the connection of the latter with the current discrepancies in cosmological measurements such as the Hubble constants or the still mysterious presence of the dark components.

therefore also simulating the space in between, in particular concentric spheres around the habited planets.

# Shutting down the simulation

The other factor to play with is the running time of the simulation.

Shorter simulations are more likely than longer ones, up to a lower threshold. The simulation threshold is the minimum (average) time after which something interesting happens. We don't know what can be considered interesting by a simulator, but we can imagine simulations interested in simulating milestones (such as development of first homo, control of fire, first communicating civilizations, first AGIs, interplanetary space travel) and then simply shutting down the simulation, taking notes of the result, and run the simulation again. Basically a simulator has no interest in simulating "boring" scenarios, that is the most likely time for the simulation to be shut off is after achieving a big milestone.

Luckily for us, this effect is not dramatic on human timescales. For instance suppose that our simulators are interested in simply simulating the control of fire, which happened a few millions of years ago. Few million years is just 0.01% of the lifetime of the universe, therefore our universe is only marginally more likely to be shut down than a universe in which our civilization just learned to control the fire.

# Conclusions

Our paper reports on some experimental consequences of the simulations argument that are not dependent on how the simulation is actually implemented. This has been possible by focusing on lower bounds and relative properties between simulations, factoring out our ignorance. We had made crucial use of the simplicity assumption, stating that simpler universes are more likely in the space of all possible simulations, which we justified heuristically. Additional justification and quantification of this assumption is perhaps the most interesting future research direction coming from this paper.

Interestingly, the connection between likely universes and our present reality give us a tool to falsify the simulation hypothesis. For instance, humanity actually developing interstellar travel and being able to expand to galactic distances would make the simulation hypothesis extremely unlikely.

In a sense, space exploration is our best bet to push the simulation argument (or the simulation itself!) to the limit.

# Acknowledgements

...

# Appendix

**An Heuristic Argument for the Simplicity Assumption**

Continuing from the main text, we know that at some point the simulators will have $c_n$ computational power per unit of time at their disposal for a time time $t_n$, with the most expensive simulation costing $c_n$ per unit of time. Without loss of generality, we can fix $c_1 = 1$ and therefore $c_n = n\, c_1 = n$. The number of all possible ways in which the simulation at any given unit of time can be performed is therefore simply the number of partitions of n, which is given by the partition function $p(n)$ defined as:

$$\sum_{n=0}^{\infty} p(n) x^n = \prod_{k=1}^{\infty} \left( \frac{1}{1-x^k} \right)$$

That for large n can be approximated as

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left( \pi \sqrt{\frac{2n}{3}} \right)$$

If we assumed that all possible simulations are sampled uniformly over many units of time, there will be on average 1 $c_n$ simulation every $p(n)$ simulations. In the same time frame, there will be $n_1$ occurrences of $c_1$. Since in general, given an integer $k < n$ there are

$$n_k = p(n-k) + p(n-2k) + p(n-3k) + \dots$$

We have

$$n_1 = \frac{1}{4(n-1)\sqrt{3}} \exp\left( \pi \sqrt{\frac{2(n-1)}{3}} \right) + \dots$$

Notice that it is very likely that $t_n$ for large n will be small with respect to small n, as technological progress will be affected by diminishing returns.

In the above we used a simplified model in which computational power grows as integer values, with $c_n$ being n times $c_1$. A more realistic case in which the simulators computational power has fractional improvements (in which for instance $c_2 = 1.1\, c_1$ ) doesn't alter our conclusions, and indeed make $c_n$ simulations even less likely, as there are many more ways of performing simpler simulations.

**Bibliography**

**[1] Bostrom, Nick. 2003. "ARE YOU LIVING IN A COMPUTER SIMULATION?" *Philosophical Quarterly (2003)* Vol. 53 (N. 211) pag. 243-255.**

[2] Bostrom, Nick. *Anthropic bias: Observation selection effects in science and philosophy*. Routledge, 2013.

[3] http://www.johnstonsarchive.net/astro/nearstar.html