# 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 measure is given by computational complexity: easier simulations are more likely to be run. 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 and that humanity is not going to meet other intelligent species in the universe, in turn explaining the Fermi's Paradox. On the opposite side, experimental falsification of any of these predictions would constitute evidence against our reality being a simulation.

*Keywords:* Simulation Argument; Simulation Hypothesis; Fermi Paradox.

# 1 The 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 computer simulations and no apparent physical limit, we are led to conclude that we will be able to build realistic civilization-size simulations, such as a computer simulation of the entire earth and human species. 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 civilization-size simulation and explore some of the observational consequences of being inside a simulation.

# 2 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. Let's begin by noticing that in the space of all possible simulations we are very likely to find ourselves among the most likely, that is the more numerous, classes of simulations. This observation follows from the so-called Self-Indication Assumption: All other things equal, an observer should reason as if they are randomly selected from the set of all possible observers [2]. In our case, the observer is an entire simulated civilization.

From our everyday experience we know that those simulations should at least be powerful enough to be able to simulate our solar system, our particle physics experiments 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 human space colonization, we can argue that simulating in detail only the entire solar system would be compatible with our observations.

Now the question is, what is the most likely class[1] of simulations compatible with the

[1] One may argue that the reference class of a civilization is not well defined. For instance is the reference class of the human species simulations including a universe in which WWII never started? Probably yes. A universe in which we evolved tails? Probably not. Without loss of generality, we can fix an arbitrary

observations that a civilization has about its own universe? An answer is given by the following assumption:

*The simplicity assumption (SA): If we randomly select the simulation of a civilization in the space of all possible simulations of that civilization that have ever been run, 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 run. Or equivalently, simpler simulations outnumber complex simulations.

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 provide two heuristic models for the probability of a simulation to be run as a function of its computational cost, showing two possible extremes respectively scaling exponentially and linearly. These two models hint that the real SA scaling is somewhere in the middle.

It doesn't matter how large, the simulators must have finite computational power at their disposal at a given time[2]. Here by simulators we mean the sum of all the entities existing at any given time that are running simulations of our civilisation, including simulated simulators. For instance these could be large international collaborations, universities, individual developers or consumers, AIs.

Suppose that at $t_{start}$ the (combined) simulators dedicate for the first time enough computational power to run a simulation of our civilization in the simplest possible way. Without loss of generality we call the associated computational power per unit of time $c_1 = 1$ and the unit of time $\tau$. The simulators allocate this simulation power for a period $t_1$, during which they limit themselves to simulations of computational cost $C_1 = c_1\tau$, for many periods $\tau$. Here we are making the restrictive assumption, which we will relax in the next model, that the simulators do not embark in simulations more complex than what they can simulate in a single period of time, using their current computational power per unit of time.

reference class and the arguments of this paper will apply to that class.

[2] By time we mean the time measured in the parent universe.

After a period $t_1$, they obtain a combined 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 $\tau$, they have the option of allocating computational resources to a single simulation of computational cost $C_2$ or to two simulations of computational cost $C_1$. Apart from an initial period in which only few entities have the resources to run these bleeding-edge simulations, for large n we have no reason to believe that allocating computationally resources in one way should be preferred, as we are considering a large number of independent simulators with different computational power and goals, some of them even in different nested levels of simulations. So we assume that, in a first order approximation, all possible ways in which the computational resources dedicated to simulations can be partitioned are sampled uniformly over many units of time. In the example above, the simulators would run $\{C_1, C_1\}$ and $\{C_2\}$ roughly with the same frequency, over many units of time $\tau$ in $t_2$.

This implies that at large $n$ for a single instance in which a simulation of cost $C_n$ is run there are also $n_1$ instances of $C_1$ simulations, where (see appendix):

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

Moreover, contrary to $C_n$ simulations, simulations costing $C_1$ were also run for all the $t_k$ with $k < n$. So $C_1$ simulations exponentially outnumber $C_n$ simulations. Using the Self-Indication Assumption, finding the occurrences of a class of simulations equals finding the probability p of being in one of those simulations, up to an unknown normalization constant $\hat{N}$:

$$p_k = \hat{N} n_k \tag{2}$$

Therefore

$$\frac{p_1}{p_n} \geq \frac{e^{\pi\sqrt{\frac{2}{3}(n-1)}}}{(n-1)} \tag{3}$$

For all the periods $t_k$ with $k \geq n$. In a nutshell, simple simulations are much more likely since for a long time they are the only possible simulations that can be run, and later they can be run a huge amount of times with respect to a limited number of times for a more costly simulation.

It's worth considering the latter statement from a different perspective: if one wanted to rule out the SA, he should provide a universal suppression mechanism for simple simulations which make simple simulations unlikely to be run as more computational power is available. While it is reasonable to imagine some simulators not being interested in running simple simulations, it's hard to imagine that all of them would not be interested. But even few simulators interested in running a large number of simple simulations can quickly outnumber all the complex simulations ever run so far with little computational effort compared to the resources available at that time.

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 run inside nested simulations, due to limits in computational power.

An example partially[3] illustrating the above points are the classic arcade video games, such as Asteroids, developed originally by Atari. Not only have these games been played billions of times, they have also been featured as playable games inside larger video games (an Asteroids clone called Duality appears in Grand Theft Auto: San Andreas) and used as test benchmarks for training reinforcement learning and artificial intelligence algorithms (the Atari 2600 benchmark [3]). Similarly, Doom (1993) can be played [4] inside Doom (2020) and it is also used as a reinforcement learning environment [5].

The previous model is overcounting the number of simulations of low complexity with respect to the high complexity ones, since the simulators would actually be able to perform complex simulations by spreading their maximum computational power over many units of time. Here we consider the opposite extreme model, in which the simulators spread their computational power uniformly among the possible $C_n$ over many units of $\tau$. More explicitly, for every $\tau$ during $t_1$ they uniformly choose between $C_1$, $\frac{1}{2}C_2$, $\frac{1}{3}C_3$, $\ldots$ , $\frac{1}{n}C_n$,

[3] Partially, since there are no sentient observers inside!

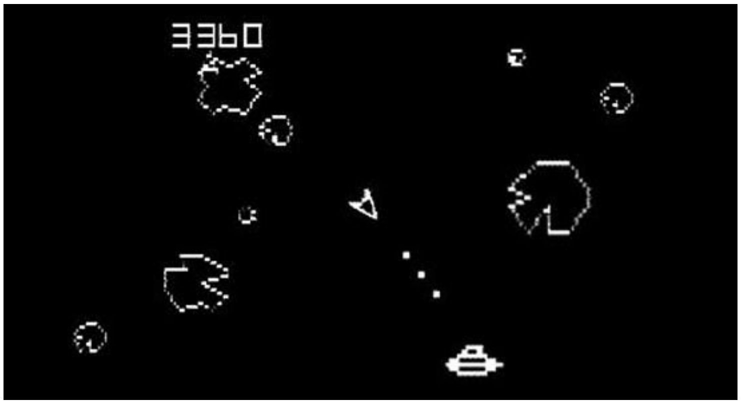

Figure 1: Asteroids (Atari-1979)

$\ldots$, $\frac{1}{2}C_1 + \frac{1}{4}C_2$ and all the other possible combinations[4], where the fraction indicates the amount of the whole simulation that is carried out.

In this model

$$\frac{n_1}{n_n} = n \tag{4}$$

so $C_1$ simulations linearly outnumber $C_n$ simulations following a (harmonic) Zipf's law behaviour and therefore

$$\frac{p_1}{p_n} = n \tag{5}$$

For all the periods $t_k$. This model is also unrealistic, but overcounting $C_n$ simulations this time, since it makes little sense for low powered simulators to start simulations that will take decades to complete, when they can simply wait for their technology to improve while they run simpler simulations. They will therefore wait to venture into $C_n$ simulations until they are close enough to have capabilities $c_n$.

Nevertheless, Zip's like models, that is power-law scaling of the rank of an instance vs its frequency, are intriguing since they model with good accuracy many human created systems, such as languages, corporation sizes, city sizes and websites traffic [6]. In particular the distribution of the computational cost of UNIX processes for different computer systems, measured as CPU age $T$, has been found to follow a power law behaviour $1/T^g$ with exponent approximately $g = 1$ [7]. The phenomenon of preferential attachment, a.k.a "the

[4] The maximum n is dictated by the limits of computation of the parent simulators universe.

rich get richer", is often invoked as a mechanism for the appearance of power laws or Yule-Simon distributions. Perhaps the simple simulations becomes more and more likely, as software and hardware to run them is increasingly optimised and standardised, so that simulators are even more enticed to run them in a cycle of self improvements.

We call the SA with linear scaling the *Weak Simplicity Assumption*, and also define the *Strong Simplicity Assumption* as the SA with exponential scaling. We can now answer the previous question: if we randomly select a simulation, using the Weak SA the simulation B is about 1000 times more likely than A to be picked, 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 by including the contribution of other kinds of complexities, for instance space computational complexity and the Kolmogorov complexity of the program running the universe (for the latter, see for instance [8]). The addition of a Kolmogorov complexity contribution looks particularly suited in the case of non-intelligent simulators, for instance in the case in which the original computer running simulations simply emerged as random fluctuation (a "Boltzmann Brain computer"). 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.

The Simplicity Assumption is somewhat conceptually related to the Speed Prior introduced in [9], which assigns $\frac{1}{n}$ probability to optimal computational processes requiring more than $O(n)$ resources.

## 3 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. In our universe 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. In the following we will consider actual atoms (hydrogen, etc.), as their number is positively correlated with the fundamental atomic entities.

The lower bound on the computational cost of the simulation is therefore proportional to three factors:

$$C \propto N\,T \propto \frac{S\,T}{A} \tag{6}$$

Where $S$ is the physical size of the high detail region of the simulation, $T$ is the simulation running time (or total life of the universe) and $A$ is the approximation level of the high detail region of the simulation, which we can think as the shortest distance over which computations take place (very detailed simulations have small $A$). Simulations with the same value of $C$ are roughly equally likely to be run.

The number of atoms to simulate a given scenario depends on the desired approximation level $A$. We can certainly imagine more detailed simulations, but what about simpler simulations of our civilization? 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. One may argue that simulators could also approximate local physics, for instance in the interior of stars and planets, but there is a fundamental difference between these two extremes: we can actively probe what is local, while we can only passively observe what is distant. The demand for local physics consistency for many active observers over long periods of time therefore put severe constraints on how much local physics can be approximated.

Anyway, the extent to which local physics is approximated doesn't qualitatively change the results of subsequent paragraphs, since given two simulations with the same $ST$ factor, the simulation with less details is more likely. Said differently, we are very likely to find ourselves in a simulation with the highest possible approximation level compatible with our observed reality at any given time. So we will treat $A$ as constant in the next paragraphs.

To summarise, the computational complexity of simulations is at least linearly proportional to the time the simulation is run, multiplied by the number of atoms. We have finally set the stage and we are ready to draw some observable consequences of living in a simulation.

# 4 Interstellar Travel

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[5]. 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 25 light years around the sun, with a total mass in the order of $10^2$ solar masses [10]. Even disregarding dark matter and interstellar gas, this means that a simulation of our universe with full rendering of the closest 25 light years is no less than 100 times more computational intensive with respect to a simulation running for the same time in which we are confined to our solar system, and therefore 100 times less likely according to the Weak SA and $10^9$ times for the Strong SA .

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 the Strong SA, the probability factor balloons to more than $10^{10^6}$) .

[5] 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 constant or the still mysterious presence of the dark components.

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).

## 5 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 aliens aren't "non-player characters", but they have similar experiences to us. Basically in the simulation no intelligent observers are preferred. Simulating far away aliens requires therefore also simulating the space in between, in particular concentric spheres around the habited planets.

A resolution of the Fermi's paradox linked to the simulation argument has been already proposed before, for instance in [11].

## 6 Shutting down the simulation

The other factor to play with is the running time of the simulation. Given the same spatial setup, 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.

It also follows that the probability of our universe having infinite lifetime is substantially zero.

## 7 So, why are we here now?

If we will never achieve interstellar travel or similar milestones, why did we arrive where we are now? After all, the SA would suggest that a universe in which humanity never lands on the moon or explores in detail the solar system is more likely than our universe. For instance earth's gravity well could have been too strong for any kind of chemical or nuclear propellant to reach critical velocity.

A possible answer is that no jump in complexity of our civilization so far is comparable to the much larger jump which we would face to be a Kardashev III interstellar species. So, our current civilisation could be not too far from the simplest possible way of simulating a civilization[6].

Finally, we should acknowledge that however unlikely our existence is, here we are. We are the only data point available. What the SA can give us, assuming that we are in a simulation, is a tool to estimate how likely we are to progress from here.

## 8 Conclusion

This 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 one of the most interesting open questions from this paper.

The connection between likely universes and our present reality, that is the SA, 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. Or one may argue that achievements we already made, such as the exploration of the solar system, are already a strong argument against the simulation hypothesis.

[6] There could also be another factor: how interesting we are. We may be not too far from the simplest way of simulating an interesting enough civilization.

Looking at the future, space exploration looks our best bet to push the simulation argument (or the computer running the simulation!) to the limit.

## Acknowledgements

...

## References

[1] Nick Bostrom. Are you living in a computer simulation? *Philosophical Quarterly*, 53(211):243–255, 2003.

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

[3] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. `https://arxiv.org/abs/1312.5602`, 2013.

[4] Playable DOOM inside Doom Eternal. `https://www.youtube.com/watch?v=VcfnTpHK28A`, 2020. [Online; accessed 1-Feb-2021].

[5] Marek Wydmuch, Michal Kempka, and Wojciech Jaskowski. Vizdoom competitions: Playing doom from pixels. *CoRR*, abs/1809.03470, 2018.

[6] Xavier Gabaix. Zipf's Law for Cities: An Explanation. *The Quarterly Journal of Economics*, 114(3):739–767, 08 1999.

[7] Mor Harchol-Balter and Allen B Downey. Exploiting process lifetime distributions for dynamic load balancing. *ACM Transactions on Computer Systems (TOCS)*, 15(3):253–285, 1997.

[8] Marcus Hutter. A complete theory of everything (will be subjective). *CoRR*, abs/0912.5434, 2009.

[9] Jürgen Schmidhuber. The speed prior: A new simplicity measure yielding near-optimal computable predictions. `https://doi.org/10.1007/3-540-45435-7_1`, 2002.

[10] Robert Johnston. List of Nearby Stars: To 25.1 light years. `http://www.johnstonsarchive.net/astro/nearstar.html`, 2018. [Online; accessed 1-Feb-2021].

[11] M.M. Cirkovic. Fermi's paradox: The last challenge for copernicanism? *Serbian Astronomical Journal*, (178):1–20, 2009.

[12] Manosij Ghosh Dastidar and Sourav Sen Gupta. Generalization of a few results in integer partitions. *arXiv preprint arXiv:1111.0094*, 2011.

# Appendix

## A Toy Model for the Strong 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 $c_n = n$. The number of all possible ways in which the simulations 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:

$$\sum_{n=0}^{\infty} p(n)x^n = \prod_{k=1}^{\infty} \left( \frac{1}{1 - x^k} \right) \tag{7}$$

That for large n (from G. H. Hardy, Ramanujan and J. V. Uspensky) can be approximated as

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left( \pi \sqrt{\frac{2n}{3}} \right) \tag{8}$$

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 unit of time, there will be $n_1$ occurrences of $c_1$ simulations. Since in general, given an integer $k < n$ one can show that [12]:

$$n_k = p(n-k) + p(n-2k) + p(n-3k) + \dots \tag{9}$$

We have

$$n_1 = \frac{1}{4(n-1)\sqrt{3}} \exp\left( \pi \sqrt{\frac{2(n-1)}{3}} \right) + \dots \tag{10}$$

In the above we used a simplified model in which computational power grows as integer values, with $c_n$ being n times $c_1$. The continuous limit doesn't alter our conclusions, and indeed make $c_n$ simulations even less likely, as there are many more ways of performing simpler simulations.