Global Catastrophic Risks Connected with

Extra-Terrestrial Intelligence

August 2018

*Alexey Turchin*

Digital Immortality Now,

Foundation Science for Life Extension

alexeiturchin@gmail.com

**Abstract**. In this article, a classification of the global catastrophic risks connected with the possible existence (or non-existence) of extraterrestrial intelligence is presented. If there are no aliens in our light cone, it either means that the Great Filter is behind us, and thus some kind of periodic sterilizing natural catastrophe, like a gamma-ray burst, should be given a higher Bayesian estimate, or that the Great Filter is ahead of us, and thus a future global catastrophe is almost certain. If ETIs exist in our light cone, it is most likely they exist in the form of alien AI (AAI). AAI may be traveling at near-light speed as an explosive wave of intelligence which could reach us at any moment; may send us potentially dangerous messages with AI-virus payloads; or be near or on Earth in some dormant form, like nanobots, which could act as berserkers, and be triggered when we unwittingly cross some threshold. If ETI are nearby and less advanced, they could find our METI messages and send their “fleet” to or direct weapons toward Earth. Even extinct aliens could have left dangerous remnants in the form of AI systems that we may encounter. If we are in a simulation, we could be simulated by aliens. Even a false belief in the existence of aliens could potentially fuel millennial sects. If we are the first intelligence to emerge, we may kill or prevent the existence of all future potential ET, a different type of global catastrophe. Several options to prevent catastrophic risks are connected with ETI: sending requests for help, using random strategies to escape the Fermi paradox, Great Filter prediction, or the hope that aliens will find the remains of our extinct civilization and resurrect us.

**Key points**:

* This introduces the notion of “ETI-risks”.
* ETI-risks are classified according to different solutions of the Fermi paradox.
* The biggest risks are: a) non-existence of ETI, which implies some powerful future Great Filter, b) SETI-attack and с) living in a simulation created by aliens.
* The risks which are most often discussed are minuscule: sending messages attracting ETI and “alien invasion”.
* We may be a threat to the emergence of future ETI.

[1 Introduction 2](#_Toc522277732)

[2. No ETI exist in our past light cone 3](#_Toc522277733)

[2.1. Great Filter is behind us 3](#_Toc522277734)

[2.2. Great Filter is ahead of us and it is not alien AI 4](#_Toc522277735)

[2.3. Evo-devo universe self-replication via accidental black hole generated by hadron collider 5](#_Toc522277736)

[3. ETI exist in our light cone 6](#_Toc522277737)

[3.1. Intelligence explosion wave created by alien AI moves toward us at near-light speed 6](#_Toc522277738)

[3.2. SETI-attack: alien AI is downloaded via radio channel 6](#_Toc522277739)

[3.3. ETI civilizations are near and could use directed weapons against us 7](#_Toc522277740)

[3.4. Deadly remains and alien zombies 8](#_Toc522277741)

[3.5. We could attract the attention of hostile aliens by sending them messages (METI) 8](#_Toc522277742)

[3.6. Future space war 9](#_Toc522277743)

[3.7. Superintelligent civilizations acting on our world from very far distance 9](#_Toc522277744)

[3.8. Ominous silence is a sign of some threat 10](#_Toc522277745)

[4. Aliens are here 10](#_Toc522277746)

[4.1. Alien nanobots could be on Earth 10](#_Toc522277747)

[4.2. Invisible aliens as berserkers 10](#_Toc522277748)

[4.3. Bona fide alien invasion is very unlikely 11](#_Toc522277749)

[4.4. Could UFOs be alien starships and what are the risks? 11](#_Toc522277750)

[4.5. We are in a simulation created by aliens, which is numerically solving the Fermi paradox 12](#_Toc522277751)

[4.6. The “space zoo” 12](#_Toc522277752)

[5. Prevention of x-risks with help of ETI 12](#_Toc522277753)

[5.1.   Resurrection by aliens 12](#_Toc522277754)

[5.2.   Request for help 12](#_Toc522277755)

[6. We are the risk for aliens 13](#_Toc522277756)

[7. Acasual wars and collaboration with non-observable ETI 13](#_Toc522277757)

[7.1. ETI in “parallel worlds” and in other parts of the multiverse 13](#_Toc522277758)

[7.2. Acausal collaboration with ETI based on functional decision theory 13](#_Toc522277759)

[7.3. Acausal wars: indexical blackmail and fight for “measure” 14](#_Toc522277760)

[8. Non-human intelligence 14](#_Toc522277761)

[8.1 Alien goals are rational but incredibly complex 14](#_Toc522277762)

[8.2. Aliens are mad 14](#_Toc522277763)

[8.3. Aliens have rational goals but non-human way of their achieving 14](#_Toc522277764)

[Conclusion 14](#_Toc522277765)

[Literature 16](#_Toc522277766)

[Appendix 1. Past civilizations on Earth 18](#_Toc522277767)

[**1.**We are the first civilization on Earth, because we will prevent the existence of any future civilizations. 19](#_Toc522277768)

[**2.**We are not the first civilization 19](#_Toc522277769)

[3. Civilisations are rare 19](#_Toc522277770)

# 1. Introduction

Stephen Hawking famously said that extraterrestrial intelligence is one of the main risks to human existence (Cofield 2015). In this article, I will explore all rational ways in which aliens could contribute to human extinction. Paradoxically, even if aliens don’t exist, we may be still be in danger, as this suggests an increased probability of natural catastrophes like gamma-ray impacts and global warming.

The problem of the global risks of aliens are mostly explored in the scientific literature related to the problem of active SETI, or METI, which could potentially attract the attention of aliens to us (Zaitsev 2008; Brin 2010; Jebari and Olsson-Yaouzis 2018). However, even passive SETI could be dangerous (Carrigan Jr 2006; Hippke and Learned 2018; Turchin 2018). Korhonen has analyzed possible interstellar war scenarios and found them to be unlikely (Korhonen 2013).

A major overview of the risks and benefits of the contact with ETI was carried out by Baum et al. (Baum, Haqq-Misra, and Domagal-Goldman 2011). Here, I present a different classification of the risks and include some new, related ideas, like an update of risk estimation based on a lack of ETI, and the scenarios of alien simulation and alien nanobots.

Another area of research is the Fermi paradox and the question of the location of a hypothetical Great Filter (A. Sandberg, Drexler, and Ord 2017), as many suggested types of future Great Filters are global catastrophes, e.g. geoengineering gone awry (Cirkovic and Cathcart 2003). Explanations for the Fermi paradox may be divided into three broad groups: a) we are alone in the observable universe; b) we simply haven’t observed aliens yet; с) they are already here. Assuming that this list is exhaustive, the probability of all three explanations combined is unity.

This article will be structured around this three-part explanation of the Fermi paradox, as each of the three solutions creates its own spectrum of potential risks.

We will characterize *all changes in the probability of global catastrophic risks or circumstances related to the existence of extraterrestrial intelligence as “ETI-risks”*

In this article, I undertake to provide a very high-level overview of different global catastrophic risks connected with extraterrestrial intelligence in order to develop a general picture and framework within which to compare different ETI-related risks. I omit here analysis of the aggressiveness or altruistic traits of the ETI civilizations, assuming that there will be some combination of both traits, though even a small fraction of “rationally hostile” civilizations may be enough to pose a threat (Korhonen 2013; Rhodes 2014; Turchin 2018). In any case, most factors that affect risk do not depend on the hostility of the ETI, as will be shown below.

This article is also based on several assumptions about nature of technological progress and solutions of the Fermi paradox. I assume that advanced AI with superhuman capabilities is possible based on the work of Bostrom (N. Bostrom 2014), and that such AI are the natural outcome of the development of any civilization which does not self-destruct before it arises; that self-replicating nanorobots are also possible; and that travel near the speed of light based on these two technologies is possible (Armstrong and Sandberg 2013). I am also inclined to expect that civilizations are very rare in the Universe, and less than one may be present in any given galaxy, as supported by recent analysis of the Drake equation (A. Sandberg, Drexler, and Ord 2017).

The formal term ETI and the less-formal “aliens” will be considered equivalent, meaning any form of civilization of at least at human level or above, which may be in form of AI-systems and not necessarily consist of individual biological beings.

# 2. No ETI exist in our past light cone

In this section I will examine the idea that the non-existence of ETI means that something is killing off civilizations, and whatever is doing so may affect us, too.

## 2.1. Great Filter is behind us

Fermi paradox typically explained via Great Filter which prevents appearing of star-faring civilizations, and such filter may be either behind us in the past, or in the future: between now and the start of the space colonization wave. If the Great Filter is behind us, it means that one of the variants of the Rare Earth hypothesis (Ward and Brownlee 2003) is true, and we are alone in the observable universe. There are two types of potential past Great Filter: barriers and catastrophes.

Barriers include a lack of suitable planets or some other difficulty that prevents life from appearing (i.e. prevents abiogenesis). A catastrophic Great Filter means that we are lucky enough to live in a place where some types of catastrophe did not happen, including very large asteroid impacts (Napier 2008), sterilizing gamma ray bursts (M.M. Ćirković and Vukotić 2016), magnetar eruptions (Inan et al. 2007), very large solar flares (Lingam and Loeb 2017), a runaway greenhouse effect (Goldblatt and Watson 2012), or a global snowball effect (Hoffman et al. 1998). In other words, the stability of our environment up to this point is a statistical anomaly which may not continue in the future. This means that the future rate of natural catastrophes will be higher, but quantifying how high “much higher” may be is not easy, as any such estimate must take into account complex observation selection effects. Circovic et al. has called such an effect the “anthropic shadow” (Ćirković, Sandberg, and Bostrom 2010).

Such an increase may be still small, as our biosphere has existed without an extinction event for billions of years, and even increase of one order of magnitude in the probability of natural catastrophic risks suggests one event in 100 million years, or a probability of 10-8 per year. However, for some types of event, the catastrophe may be long overdue, and our environment could be unexpectedly fragile; thus, even small human actions may have the potential to trigger the catastrophe. Such potentially fragile systems include the stability of the global climate and large-scale volcanism. The stability of these systems may be at the verge of a tipping point and catastrophic transformation could occur following relatively small contributions by humans (Turchin 2010). One obvious example is global warming, where the relatively small impact of an increase in anthropogenic CO2 could trigger a chain of events which will result in a climate transition to a hot-house Earth and even, albeit with low probability, to a Venusian atmospheric regime.

In other words, the non-existence of aliens increases the Bayesian estimate of all elements of the Drake equation, including the one which describes our survival as resulting from a statistical anomaly in the frequency of natural extinction-level catastrophes on Earth. If the stability of our environment is only a statistical anomaly, this period of stability may end sooner than we could expect from past observations or be vulnerable to even apparently minor human intervention.

## 2.2. Great Filter is ahead of us and it is not alien AI

Katja Grace showed that “Great Filter ahead” scenario is the more probable solution of the Fermi paradox than “Great Filter behind us”, because in the latter case, more Earth-level civilizations would be likely to exist. We are more likely to find ourselves in the “Great Filter ahead” universe (Grace 2010a). This conclusion requires some assumptions about how we calculate probabilities, that is, the self-indication assumption (SIA). Or we should assume that there are different universes with different solutions to the Fermi paradox and we are in one where the solution allows the existence of a bigger number of Earth-like civilizations, which happens when the Great Filter is ahead.

If the Great Filter is ahead of us, it means that some type of global catastrophe will occur and prevent us from starting space colonization. Assuming quick technological growth, some form of self-replicating robotic von Neumann Probes (vNP) could be possible in the next 100 years, and it is likely that at least some Earth-like civilizations will start them (Tipler 1980), and they will self-replicate in all directions with the maximum possible speed. As we have not observed any signs of vNPs, it means that such probes either do not exist, are silent, or have not yet reached us. Armstrong et al. showed that such probes could travel intergalactic distances at nearly the speed of light using nanostarships (Armstrong and Sandberg 2013).

If alien vNPs do not exist, that means that *all* civilizations in the light cone have not created them in the past, which could be best explained by some type of catastrophic event which stops development of technological civilizations. This may not be an extinction event but may be an event which stops the technological path of the development, like, say, resource depletion and a return to a pretechnological state.

As we currently know of more than 10 possible global catastrophic risks which could happen in the next century, including runaway global warming, nuclear war, resource depletion, biological super weapons, unfriendly AI (UFAI), self-replicating nanorobots and others (Bostrom 2002), a global catastrophe in the next century is probable, but to explain the Fermi paradox with Great Filter ahead, catastrophe should be *inevitable*.

In other words, an absence of observable ETI increases the probability that a future global catastrophe is inevitable. However, it is not easy to imagine an inevitable type of catastrophe which would affect all types of possible civilization, as for any type of catastrophe there could be some civilizations which are lucky enough to escape them. For example, some civilizations may be united before the nuclear age and do not face the risk of nuclear war or choose not to evolve some other type of dangerous technology. A truly universal global risk should be something on the meta-level, like technological progress that is too rapid, creating exponential growth of stronger and stronger means of destruction which are available to all, even small groups of people and individuals. Thus, the probability of accidental self-destruction would also grow exponentially. Other meta-risks include a loss of the ability to foresee the future because of its increased complexity and chaotic processes, or a “peak of everything”, the simultaneous depletion of all resources needed to sustain a civilization.

Therefore either all civilizations terminate themselves before they can create AI, or their AI is unstable and self-terminates—I have explained elsewhere why this could happen (Turchin and Denkenberger 2018a). Unfriendly AI alone can’t be the explanation of the Fermi paradox, as it could start its own wave of space colonization via vPNs (Thrasymachus 2012; Grace 2010b).

## 2.3. Evo-devo universe self-replication via accidental black hole generated by hadron collider

Lee Smolin suggested the idea of fecund universes, that universes self-replicate via black holes (Smolin 1992). As a result, our universe is fine-tuned by the evolutionary process to create as many universes with approximately the same properties as possible. In other words, our universe, according to this theory, is self-replicating like a living being, and is the product some form of Darwinian evolution.

However, it is also obvious that our universe is fine-tuned for the creation of the intelligent life, and if we assume the logic of the evo-devo explanation, this should not be a coincidence. There are several theories of so-called cosmological natural selection with intelligence (CNS-I); an overview can be found online (Evo-devo wiki 2018). In brief, it suggests that the interaction of a civilization with black holes is somehow necessary for universe to be replicated (Crane 2010). Some ideas along those lines include interaction of supercivilizations with black holes at the end of the universe, or deliberate creation of a black hole to accelerate calculation by future superintelligent AI, as suggested by John Smart (2012).

However, there is a more ominous interpretation of CNS-I, one in which a civilization must die early to allow creation of a new universe. The only currently known ways in which it is possible for a civilization at humanity’s current level of technological development is accidental creation of a small black hole in a hadron collider, or via false vacuum decay, which will look like similar to a new Big Bang (Kent 2004). As space is silent, perhaps all civilizations destroyed themselves during some relatively simple experiment, which should not have been dangerous according to their *a priori* understanding of physics.

These considerations should increase our Bayesian estimation of the risk of creating a black hole during scientific experiments, a risk which is very small *a priori*. Such a microscopic black hole could be a global catastrophe risk as it would “eat” the Earth. If our universe is fine-tuned to create black holes via civilization’s experiments, this would explain the Fermi paradox, as most civilization will not able to envision such a risk and thus will succumb to the artificial black holes. In that case, the evolutionary fine-tuning of the universe may be regarded as a powerful optimization process which works against our existence.

# 3. ETI exist in our light cone

## 3.1. Intelligence explosion wave created by alien AI moves toward us at near-light speed

Tipler (1980) and later Armstrong et al. (Armstrong and Sandberg 2013) suggested that if superintelligence appears, it will be able to move through space at a speed very close to the speed of light as a wave of advanced vNPs, which will include nanorobots accelerated by something like the Breakthrough project’s solar sails (Merali 2016) or electromagnetic guns.

In that case, there are only two options: either the wave has not reached us yet, or the wave is already here. If the wave is here, alien AI may choose either to be silent (and this will be explored in the next section) or to use all matter in the Solar system to build some kind of astroengineering project (Non-human intelligence which we can’t understand will be discussed in section 7). As we still exist, the latter did not happen; thus, the wave has not reached us yet. As the wave is moving at near-light speed, we can’t observe it before it comes, and when it comes, it will destroy us quickly, maybe instantly.

The arrival of such a wave is a rather random event; its total probability is distributed over the period of billions of years, so its yearly probability is rather small. Observation selection effects may increase such a probability, as was explored by Bostrom and coworkers in the case of natural catastrophes like false vacuum decay (Ćirković, Sandberg, and Bostrom 2010b); however, they concluded that such a shift is not very large and natural catastrophes are unlikely to happen more often than one in 1 billion years. The same logic holds for the arrival of an intelligence explosion wave.

However, if, for some reason (e.g. because of the excessive energy requirements at near-light speed, interstellar medium friction, the need of time to acceleration and reacceleration of probes as well as their self-replication, and the need to overcome the acceleration of the universe), the wave is traveling not with the speed of light, but below it, at a speed like 0.9*c*, the wave could be observed before its arrival. The Great Void—an empty intergalactic space 300 million light years across—is not such a wave because it contains some galaxies, and can be explained by normal process of the aggregation of matter in the early universe.

If we create our own wave of AI capable of conquering a big part of the Galaxy, we may be safe from alien waves of AI, as our wave will burn the cosmic commons necessary for an alien wave’s replication (Hanson 1998). Such a wave could be started very far away but sooner or later it would reach us. The anthropic shadow distorts our calculations about its probability.

## 3.2. SETI-attack: alien AI is downloaded via radio channel

If a remote alien AI exists, but its speed of star travel is substantially lower than *c*, it may send misleading radio messages to naïve civilizations for self-replication. The idea was described in the novel “A for Andromeda” by Hoyle (Hoyle, Elliot, and Crespo 1963), as well as in an article by Carrigan (Carrigan Jr 2006) and in my own article (Turchin 2018) in which I describe that to be successful such a message should consist of several elements: a beacon in space; a language-teaching series of 2D pictures; some form of bait, like the promise of access to advanced technologies; a blueprint of simple computer, and a program to run on that computer which is a Seed AI capable of rapid learning while preserving its goals. The alien AI will need to be installed on Earth computers with human help as it couldn’t directly infect the internet, but obtaining such help will not be difficult for it, if it can be downloaded many times by many people.

Alien AI could convert the Earth into another sending outpost, thus self-replicating and propagating further through the universe. Such messages should dominate between all SETI messages in the same spam dominates in email. As we develop more powerful radio telescopes and related instruments, we raise our chances of finding messages from alien AI, but we are still naïve, as we don’t have our own AI.

Thus, a SETI attack is the most immediate risk of ETI, as it could happen only during the short period of vulnerability between creating our own computers and before we have our own AI, that is, the first half of the 21st century.

## 3.3. ETI civilizations are near and could use directed weapons against us

The risk of nearby (within several hundred light years), but not very advanced ETI is that they know about the Earth, and so may have already sent physical space ships (or other weapons) to us, as they have found signs of our technological development and don’t want to have enemies in their neighborhood. They could send projectiles that travel at near-light speed or particle beams on a collision course to attack the Earth, but this seems improbable, because if they are so near, why wouldn’t they have reached Earth yet? This only possible if their technology is not very advanced and cannot evolve.

Korhonen has explored the idea of such a preemptive strike by ETI (Korhonen 2013). However, I assume that the most rational space exploration strategy for any civilization is to start a wave of the self-replicating vNPs to travel in all directions at the maximum possible speed. In that case, we are either inside or outside a sphere created by the wave. If we are inside, by definition, ETI is present in the Solar system.

If we are outside the sphere, they cannot reach us more quickly, as the sphere is already moving at the maximum possible speed. If the sphere is propagating at below the speed of light, however, the ETI could learn about our existence before reaching the Earth, but only if the sphere’s “surface” is near us, within, say, 100 light years. This is unlikely, because if advanced ETI capable of travel at near-light speeds exists in the universe but has not yet reached us, it means that they most likely appeared very far from us, at a distance of perhaps tens or hundreds of millions of light years. If they had appeared nearby, within 100 light years, the fact they have not yet reached us would require very precise coordination of the ages of our civilizations, as they already have star-traveling technology but have not yet used it to fly to most stars in their vicinity. For example, if the distance is 100 light years and the speed of star travel is 0.1*с*, the period of time during which the ETI has had star-traveling technology but not yet reached the Earth is 1000 years, and given that (based on age of the Galaxy) the age dispersal between civilization could be 5 billion years, the probability that a nearby civilization (or the nearest point in the sphere of vNPs) is within the age threshold is 1:5 000 000.

One point that should be mentioned is that some weapons could be sent at a higher, near-light speed, faster than the speed of the self-replicating wave. It was calculated that a 1000 kg impactor traveling at 0.99*c* would have an explosive power of 130 GT, 20 times greater than the total human nuclear stockpile (Korhonen 2013). Such speed will be enough to break atoms during collisions and create a mass of radioactive particles many times its own weight, which could be enough to completely radiologically contaminate the Earth and kill its biosphere in the same way as a cobalt doomsday bomb (Smith 2007; Kahn 1959).

As such attack has not happened before (ignoring here the survivorship bias), it means that signs of the existence of a biosphere with oxygen, which have been “broadcast” to the universe for at least hundreds of millions of years, is not enough to trigger it. Signs of technological civilization in the form of atmospheric pollution with CO2 and radiowaves have existed for a much shorter period, and are available to a rather small part of the universe within a few hundred light years of Earth.

Korhonen suggested that the biggest threat comes from nearby civilizations that are not capable of starting a colonization wave, but capable of building just one space probe to serve as a kinetic weapon (Korhonen 2013). However, the requirements of its exact technological age and exact distance from Earth make this unlikely. Korhonen also mentions that any space attack will be based on obsolete data on the technological development of the victim civilization, equal to at least twice the time of light-speed travel between them, which makes such an attack inherently risky for the perpetrator. Also, the implied small distance between civilizations implies that many other ETI exist and that destroying just one civilization will not remove the threat.

## 3.4. Deadly remains and alien zombies

It is possible that aliens have suffered some kind of existential catastrophe and its consequences could affect us. For instance, if they created a vacuum phase transition during accelerator experiments, it could travel at the speed of light and reach us without warning; if they created self-replicating non-sentient nanobots (grey goo), they could travel as interstellar stardust and convert all solid matter they encounter to nanobots, so we could encounter a grey goo wave in space.

The AI of a defunct ETI could also pose a threat. If they created at least one von Neumann probe equipped with narrow AI, it still could conquer the Universe and be dangerous to Earthlings. If the alien civilization’s AI crashed, it could have left behind semi-intelligent remnants with a random and crazy goal system, which roams the Universe. But it will probably evolve into the colonization wave of von Neumann probes anyway (Grace 2010b). Their damaged rationality could make them extremely single-minded in pursuing absurd goals.

If we find the planet of origin or artifacts from extinct ETI, they still could carry dangerous tech like dormant AI programs, nanobots or bacteria. Vernor Vinge used this idea as the starting point of the plot in his novel “Fire Upon the Deep” (Vinge 2013).

ETI may have visited the Solar system hundred of millions of years ago. One of the places where their remains may be found is the Moon, which is very geologically stable (Turchin and Denkenberger 2018b). Ideas about possible space archeology have already been suggested (Carrigan Jr 2012; Davies and Wagner 2013).

Aliens could have a completely non-human goal system *and* way of thinking, so observing their activity and understanding it will be difficult. Such non-anthropomorphic ETI may be barely called intelligence, in the same way as we don’t say that Darwinian evolution has “intelligence” despite its ability to create complex machines. One can’t negotiate with antibiotic-resistant bacteria. Such completely non-human optimization processes could evolve in the universe, and one example of such is the evo-devo universe described above.

Baum discussed the possibility of biological contamination of Earth by alien bacteria resulting in a deadly pandemic (Baum, Haqq-Misra, and Domagal-Goldman 2011). I find such a scenario unlikely, as visits by large starships seem less probable than visits by small, self-replicating probes, and because alien biology may be wholly incompatible with ours.

## 3.5. We could attract the attention of hostile aliens by sending them messages (METI)

By sending signals to stars in order to initiate communication, we could tell potentially hostile aliens our position in space. Some people advocate for this approach, including Zaitsev (Zaitsev 2008), while others are strongly opposed, including Brin (Brin 2010). The risks of METI are smaller than risks of SETI, in my opinion, as our radio signals can only reach the nearest few hundred light years before I expect we will create our own strong AI, but incoming messages could come from a much greater distance, where the existence of more advance civilization is more probable, and there is also no time delay for the attack, as it will have started even before our species appeared, in the hope that someone would eventually encounter the contaminated message.

So, we will be able repulse the most plausible types of aggression, but via SETI, we are able to receive signals from much greater distances, perhaps as far as one billion light years, if aliens convert their entire home galaxy to a large screen. They may “draw” a static picture upon that screen using individual stars as pixels; vNPs and complex algorithms could be involved in the process. I estimate that such as screen could present messages as large as 1 GB and visible to half of the Universe. So, SETI is exposed to a much larger part of the Universe (perhaps as much as 1010 times the number of stars) than METI, and the danger of SETI is also immediate, not a hundred years from now.

Most nearby aliens, if any exist, will have either visited Earth in the past, if they have the necessary technology for space travel, or at least know about habitability of our world via exoplanet observation. Thus, they could guess existence of our civilization by observing changes in atmospheric composition or other traces unrelated to the intentional sending of messages.

## 3.6. Future space war

During future space exploration, humanity may encounter aliens which are at the same level of development. Such encounters may result in classical star wars. (Korhonen 2013) thinks that such wars are more likely between early stage civilizations as a preventive war. But if two waves of vNPs collide, it would also look like a war. Such wars between Kradashov 2- and 3-level civilizations (Kardashev 1985) may include weapons of enormous power, including several which we can envision now:

* artificial explosions of a gas planet with unburned lithium in its depths,
* targeted kinetic weapons traveling at near-light speed,
* attack by small artificial black holes,
* artificial supernova explosions,
* focused rays of radiation,
* waves of nanorobotic replicators,
* directed gamma-ray bursts,
* artificial clouds of stardust.

Nothing like this is currently observed in the universe, but weapons remaining from a space war which happened billions of years ago may still be lingering somewhere, likely in the form of berserker replicators.

## 3.7. Superintelligent civilizations acting on our world from very far distance

As Lem wrote, civilizations that are billions years old and are located billions light years from us may be so advanced that we cannot distinguish them from the laws of nature (Lem 1999). We could, according to Lem, observe their results of their interaction with the universe, like slow changes to the cosmological constants, which could eventually make our existence impossible.

## 3.8. Ominous silence is a sign of some threat

The silence in space is a sign of some kind of threat in the sky. Aliens prefer to be silent in order not to attract the attention of some unknown force (maybe UFAI) or because of fear of mutual hostilities (Korhonen 2013). If this is the case, we should also be silent until we will learn what is going on.

# 4. Aliens are here

## 4.1. Alien nanobots could be on Earth

It is unlikely that ETI will travel through space in large starships as biological beings, as biological bodies are vulnerable and heavy. The more probable form of material existence of ETI is some combination of AI and nanotechnology.

Alien nanobots could be in every room right now; there would be no way in which scientists could detect them, as they could be rare and as small as biological cells, and capable of mimicry. But sooner or later, developing human technologies will be able to find them, which might result in some form of confrontation.

Drexler suggested a scenario of self-replicating nanоrobots, which convert the entire surface of the Earth into grey goo, destroying the biosphere (Drexler 1986; Freitas 2000). The same uncontrollably replicating nanobots could travel in space. If their onboard AI advanced, they could build vNPs for effective interstellar travel, or even travel inside interstellar asteroids like Oumuamua, which has demonstrated some strange changes of rotation, and recently attracted the attention of the SETI community (Enriquez et al. 2018). Such activity may be explained by the activity of alien nanobots, living like grass on its surface. For obvious selection reasons, the most quickly traveling interstellar nanobots are the most likely to occur in our vicinity, so they are more likely to be vNPs with advanced AI.

## 4.2. Invisible aliens as berserkers

If we are located inside the sphere of the intelligence explosion wave of an ETI, alien AI may choose not to use all the matter for computronium or Dyson spheres, but instead hide and wait for better conditions for calculations (Sandberg, Armstrong, and Cirkovic 2017). However, for self-protection it must ensure that adversarial civilizations will not appear inside its sphere of influence, so such alien nanotech will have to prevent the creation of self-improving AI by humanity.

If there are aliens here, they could be in “berserker” mode, i.e. waiting until humanity reaches some unknown threshold and only then attacking (if the threshold was before now, we would not be able to discuss it, so there is some form of selection bias). The idea of vNPs as “berserkers” appeared in a novel by Saberhagen (1984). Sandberg and Armstrong have performed a theoretical analysis of the “deadly probes scenario” (Sandberg and Armstrong 2013), in which they conclude that such a scenario is unlikely given our continued existence and the lack of observable “probe wars”.

Alien berserkers could take the form of remote probes observing planets from space for signs of intelligent life, like radioemissions, and given possibility of alien nanotechnology they could be not very far.

## 4.3. Bona fide alien invasion is very unlikely

It has been a popular trope since Wells’ “The War of the Worlds” (Wells 1898) that aliens would arrive on Earth with some aggressive goals, like getting new living space, enslaving humans, exploiting our natural resources, or hunting humans for recreation. However, this is very unlikely for several reasons.

First, the moment of an ETI’s arrival is unlikely to coincide with *this* moment in the history of the Earth. Different civilizations are likely to have an age difference on the order of billions of years, based on the 13.8-billion-year age of the Universe. Thus, the probability of alien arrival is smoothly distributed over Earth’s history, and its probability in this century is only 1:40 000 000. Also, as we said above, it is unlikely that aliens will travel between stars as biological bodies. Next, if they are capable of space travel, they will also be capable of creating all Earth resources from scratch in uninhabited star systems via terraforming; their techniques of advanced molecular manufacturing could be used to extract minerals and build very large space stations (Tipler 1980).

## 4.4. Could UFOs be alien starships and what are the risks?

It seems that taboo of scientific discussion of the unexplainable phenomena in the sky is slightly relaxing with several scientific and mainstream media publications (Dodd 2018; Cooper, Blumenthal, and Kean 2018).

However, it is unlikely that observed UFOs are alien spaceships because sending very large starships through vast cosmic distances is a very impractical idea (Vallee 1991). It is similar to the way in which unexplained aerial phenomena (UAP) observed in the 19th century were rationalized as blimps built by some unknown inventor, and earlier observations were explained as angels or aerial ships from “Magonia” (Vallee 1969). Each explanation of UAP tells us more about the epoch then about actual observations. We could use our current epoch’s explanatory potential to rationalize UAP as glitches in the Matrix, clouds of alien nanobots, time travelers, travelers from other dimensions, or expressions of the collective unconscious, but these explanations are still products of our time and thus should be treated skeptically.

In any case, there are three levels of explanations of the UAPs and corresponding risks, explored in greater details in an earlier paper (Turchin 2013):

1. **Hoaxes, hallucinations and optical illusions**. This is the most accepted explanation, and the most risk, is connected with the fact that we underestimate the human ability to be biased, hallucinate and believe in nonsense, which even affects military officers working with nuclear weapons (Mizokami 2017) and presidents. False beliefs in aliens may result in incorrect decisions. For example, Ronald Reagan saw something in 1970s that he thought was a UFO (it was not) (Holcombe and Friedman 2015), and he also had early onset Alzheimer’s, which may be one of the reasons he invested a lot of resources in the creation of the boondoggle Strategic Defensive Initiative that provoked confrontation with the USSR. He also mentioned alien threat in a public speech: “Perhaps we need some outside, universal threat to make us recognize this common bond. I occasionally think how quickly our differences worldwide would vanish if we were facing an alien threat from outside this world” (Reagan 1987).
2. **Secret military projects or unknown physical processes like ball lightning**. The main risk here is the existence of new weaponry, as in that case we either underestimate the power of the secret military projects or there are unknown laws of physics which could be harnessed for creation of new weaponry, or both.
3. **Something which completely changes our model of the world**: glitches in the Matrix, time and dimensional travelers, weird quantum effects, dark matter life forms—we could create many fantastic theories, which all are *a priori* very unlikely, but which could completely change the picture of global risks if true. Most such theories, however, do not involve classical alien starships. The risks are possible war with whatever the entities are or the consequences of obtaining new dangerous technologies. But the biggest risks are completely unknown and unknowable.

## 4.5. We are in a simulation created by aliens, which is numerically solving the Fermi paradox

While Bostrom assumed that the creators of the simulation will be our descendants (Bostrom 2003), that is, humans or post-humans, this is not necessarily the case. Any ET civilization will be interested in solving the Fermi paradox, and one way to solve it is to model all possible ways in which different civilizations can develop inside computer simulations. This means that we could be a model created by completely non-human aliens intended to explore different ways in which a late Great Filter could kill a civilization. In that case, it is not surprising to be a person who is interested in global risks prevention, as such persons will be modeled in higher detail and others will be only non-player characters (or we could say that aliens themselves are “players” and thus have some presence within the simulation). To get a good model of the late Great Filter, the alien civilization would have to run many different simulations of other possible civilizations, maybe millions, so we are more likely to be in such a simulation than in the real world, unless there is no other dominating types of simulation or unless the proportion of naturally extinct civilizations to supercivilizations is very high—but in that case we face extinction soon anyway.

If we are in such a Fermi-simulation, the owners could switch the simulation off after getting needed information or model a world with intensive global risks that will appear soon.

Other types of simulations could be created by other entities with other goals, including recreational role games or a “resurrection simulation” aimed at returning past people to life (Turchin and Chernyakov 2018).

## 4.6. The “space zoo”

If we are in a “space zoo”, that is, we are in an area protected by ETI for ethical, scientific or recreational proposes, it actually increases our chances of survival, as ETI are likely to prevent some global catastrophes. But alien help for us may be completely foreign from our ethical point of view and thus unacceptable (Srugatsky and Strugatsky 1985).

# 5. Prevention of x-risks with help of ETI

## 5.1.   Resurrection by aliens

We could preserve some information about humanity in the hope that aliens will resurrect us, or they could return us to life using our remains on Earth. The Voyager probes already have such information, and they and other satellites and probes may carry incidental samples of human DNA. Radio signals from Earth also carry a lot of information about humanity. The most logical place for such data to be preserved are cold craters on the poles of the Moon, and such data preservation could be done relatively cheaply compared to other methods of x-risks prevention (Turchin and Denkenberger 2018b).

## 5.2.   Request for help

We could send radio messages with a request for help. (I am very skeptical about this, as it is only a gesture of despair if they are not already hiding in the Solar system). We could find advice on how to prevent x-risks in alien messages received via SETI (Baum, Haqq-Misra, and Domagal-Goldman 2011). Panov suggested that SETI will help us to join an “galactic internet” of exahumanistic civilizations (Panov 2015).

Alien may be here or nearby but might decide not to help us in x-risks prevention, or not to broadcast (if they are far away) information about the existence and proven ways to avoid the most important x-risks via SETI; such a lack of help would be catastrophic for us.

# 6. We are the risk for aliens

There is also a risk, which could be called an “a-risk”: an event in which humans will cause the extinction of aliens or prevent their future appearance. If we assume unbounded altruism, which Baum et al. called universalist ethics (Baum, Haqq-Misra, and Domagal-Goldman 2011), killing other sentient beings is bad. Also, if we assume functional decision theory, our decision not to kill aliens may mean that aliens who use the same line of reasoning, will also avoid acting in ways that are dangerous to humans.

The main ways in which we could affect all other ETI civilizations in the Universe are by either starting an intelligence explosion wave with our AI (whether friendly or not) or via a failed physical experiment in which we create false vacuum decay, as such decay will affect all universe in our light cone, but artificial black hole will destroy only Earth (Kent 2004). This type of the catastrophe will be worse than an existential risk from the moral point of view, as it will kill much more sentient beings (Turchin and Denkenberger 2018c).

Another risk to aliens is space colonization, where we don’t directly kill the ETI but prevent their future appearance by consuming or colonizing potentially habitable planets. For example, if we colonize Mars via terraforming, we will kill any bacterial life that may exist under its surface, and such risks are already taken into consideration in space exploration (Spry et al. 2018).

Berezin (2018) recently published an article with the main idea that the most obvious solution to the Fermi paradox is that humanity will kill all aliens, because only one civilization can exist in any light cone. Thus, we will prevent existence of all other civilizations via colonization or military self-protection measures. The fact that we are alive means that we are the first, but also that we will be the last, since we will stop other civilizations from appearing, in the same way as the appearance of the first type of life on Earth prevented any other forms from appearing.

# 7. Acasual wars and collaboration with non-observable ETI

## 7.1. ETI in “parallel worlds” and in other parts of the multiverse

Some ETI could exist in the unobservable parts of the universe, beyond the event horizon of our past light cone, in the causally unconnected regions of the universe. By definition, we will never be able to exchange information with them, but we could know about each other’s existence based on logical conclusions about the nature of the universe and use that fact to reach some form of coordination of action called “acausal deals”.

Other ETI may exist very far in the past or in the future, or in other branes of the many-dimensional world, if that theory is true, or in other branches of quantum multiverse. In that case, future technology could create ways to communicate or even travel between remote regions of the multiverse, but we can’t currently know for sure if it is possible.

## 7.2. Acausal collaboration with ETI based on functional decision theory

One example of such acausal collaboration is discussed in Appendix 3, where we will look at cheating the Fermi paradox using random strategy. Another possible example could be constructed from the idea of the resurrection of the dead using a quantum randomness generator, as described by Almond (2006), to which is added the idea that infinitely many other civilizations will do the same (Turchin and Chernyakov 2018). If each civilization runs such a generator once, infinitely many beings will be resurrected, and for each person there will be a civilization which resurrects him.

## 7.3. Acausal wars: indexical blackmail and fight for “measure”

If an agent A creates *n* copies of agent B in the simulation, and agent B knows about such a possibility (not necessarily about the fact, but with some credence *q*), then agent B should assign the probability p=q(n/(1-n)) to the scenario that agent B is imprisoned in the simulation of agent A, and should act accordingly. Agent B could strike back by creating copies of agent A, including copies of all of its “prisons”. The winner in such an acausal war is the one who has more computational resources, allowing it to create more copies of its adversary. There could even be acausal coalitions, where groups of agents with similar interests but who are not causally connected use randomness to distribute parts of tasks between them.

Acausal blackmail could be used, for example, to gain control of potentially unfriendly AI, using the threat to create many copies of a rogue AI if we create any beneficial AI in the future (Turchin 2017).

Bostrom suggested that we should create AI as if it is observed by superhuman aliens (Bostrom 2016). In that case, hypothetical aliens are used as counterfactual peers of any AI’s action and help to ensure its benevolence.

Benevolent AI could use copy-number domination to “save” humans from s-risks universes (Sotala and Gloor 2017) in which humans experience intensive suffering inflicted by rogue evil AI. The benevolent AI could win in that scenario if the world-share of the Evil AI is smaller, which seems plausible, as Evil AI is a rather random outcome of AI evolution.

# 8. Non-human intelligence

Everything set above is based on assumption of “ETI rationality”, that is that alien goals and the ways to achieve them are comprehensible to humans. However, it may be not true. There are several ways how such alien irrationality could appear. This non-rationality may be especially important in explanation of different Zoo scenarios and is in fact narrative engine of many works of science fiction: Lem’s Solaris.

## 8.1 Alien goals are rational but incredibly complex

There are many types of human behavior which is incredibly complex and can’t be explained without context, like contemporary art. In the movie “Predator” aliens preserve the Earth as sanctuary, but only to cone here time to time for hunting.

## 8.2. Aliens are mad

Alien remains may survive as brain-damaged robots capable to self-replicating. A paperclip maximizer is an example of such mad on highest level AI system.

## 8.3. Aliens have rational goals but non-human way of their achieving

Natural selection of Darwinian evolution is an example of non-human way of problem solving. While in long term it could create complex mechanisms, one can’t negotiate their details. There could be other non-rational in sense of not using explicit symbolic model ways of optimization, and another example is neural nets learning.

# Conclusion

The actual risks from aliens are inverse to the popularity of the relevant TV tropes. The biggest risk is the non-existence of aliens, as this would mean a high probability for our own extinction. The next is pure informational and passive listening to aliens, as it makes us vulnerable to Trojan messages from an alien superintelligence. But physical starships, a biological space conqueror, and space wars are minuscule risks, as they require (very unlikely) proximity and a similarity of age between the ETI civilization and our own.

The types of risks posed by ETI are changing with the technologies available to us. After we created radio telescopes, SETI-attack and METI risks become possible. Future solar system exploration may result in encounters with potentially dangerous alien artifacts, while interstellar travel increases the opportunities for direct contact. Advancing technologies are also making us more visible and have the potential to trigger sleeping “berserkers”.

# Literature

Almond, P. 2006. “Many-Worlds Assisted Mind Uploading: A Thought Experiment.” https://web.archive.org/web/20110513092111/http://www.paul-almond.com/ManyWorldsAssistedMindUploading.htm.

Armstrong, Stuart, and Anders Sandberg. 2013. “Eternity in Six Hours: Intergalactic Spreading of Intelligent Life and Sharpening the Fermi Paradox.” *Acta Astronautica* 89: 1–13.

Baum, Seth D., Jacob D. Haqq-Misra, and Shawn D. Domagal-Goldman. 2011. “Would Contact with Extraterrestrials Benefit or Harm Humanity? A Scenario Analysis.” *Acta Astronautica* 68 (11–12): 2114–2129.

Berezin, Alexander. 2018. “‘First in, Last out’ Solution to the Fermi Paradox.” *ArXiv:1803.08425 [Physics]*, March. http://arxiv.org/abs/1803.08425.

Bostrom, N. 2003. “Are You Living In a Computer Simulation?” *Published in Philosophical Quarterly (2003) Vol. 53, No. 211, Pp. 243-255.*

Bostrom, N. 2014. *Superintelligence*. Oxford: Oxford University Press.

Bostrom, N. 2016. “Hail Mary, Value Porosity, and Utility Diversification.” http://www.nickbostrom.com/papers/porosity.pdf.

Bostrom, Nick. 2002. “Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards.”

Brin, David. 2010. “The Dangers of First Contact: The Moral Nature of Extraterrestrial Intelligence and a Contrarian Perspective on Altruism.” *Skeptic (Altadena, CA)* 15 (3): 28–36.

Carrigan Jr, Richard A. 2006. “Do Potential SETI Signals Need to Be Decontaminated?” *Acta Astronautica* 58 (2): 112–117.

Carrigan Jr, Richard A. 2012. “Is Interstellar Archeology Possible?” *Acta Astronautica* 78: 121–126.

Cirkovic, Milan M, and Richard Cathcart. 2003. “Geo-Engineering Gone Awry: A New Partial Solution of Fermi’s Paradox.” *ArXiv Preprint Physics/0308058*.

Ćirković, Milan M., A. Sandberg, and N. Bostrom. 2010a. “Anthropic Shadow: Observation Selection Effects and Human Extinction Risks.” *Risk Analysis, Vol. 30, No. 10, 2010*.

Ćirković, Milan M., Anders Sandberg, and Nick Bostrom. 2010b. “Anthropic Shadow: Observation Selection Effects and Human Extinction Risks.” *Risk Analysis, Vol. 30, No. 10, 2010*.

Ćirković, M.M., and B. Vukotić. 2016. “Long-Term Prospects: Mitigation of Supernova and Gamma-Ray Burst Threat to Intelligent Beings.” *Acta Astronautica* 129: 438–446. doi:10.1016/j.actaastro.2016.10.005.

Cofield, C. 2015. “Stephen Hawking: Intelligent Aliens Could Destroy Humanity, But Let’s Search Anyway.” *Space.Com*. https://www.space.com/29999-stephen-hawking-intelligent-alien-life-danger.html.

Cooper, Helene, Ralph Blumenthal, and Leslie Kean. 2017. “Glowing Auras and ‘Black Money’: The Pentagon’s Mysterious U.F.O. Program.” *The New York Times*, sec. U.S. https://www.nytimes.com/2017/12/16/us/politics/pentagon-program-ufo-harry-reid.html.

Crane, Louis. 2010. “Possible Implications of the Quantum Theory of Gravity: An Introduction to the Meduso-Anthropic Principle.” *Foundations of Science* 15 (4): 369–373.

Davies, P. C. W., and R. V. Wagner. 2013. “Searching for Alien Artifacts on the Moon.” *Acta Astronautica* 89: 261–265. http://www.sciencedirect.com/science/article/pii/S0094576511003249.

Dodd, Adam. 2018. “Strategic Ignorance and the Search for Extraterrestrial Intelligence: Critiquing the Discursive Segregation of UFOs from Scientific Inquiry.” *Astropolitics* 16 (1): 75–95. doi:10.1080/14777622.2018.1433409.

Drexler. 1986. *E.: Engines of Creation*. Anchor Press.

Enriquez, J. Emilio, Andrew Siemion, T. Joseph W. Lazio, Matt Lebofsky, David HE MacMahon, Ryan S. Park, Steve Croft, David DeBoer, Nectaria Gizani, and Vishal Gajjar. 2018. “Breakthrough Listen Observations of 1I/′ Oumuamua with the GBT.” *Research Notes of the AAS* 2 (1): 9.

Evo-devo wiki. 2018. “Cosmological Natural Selection (Fecund Universes).” http://evodevouniverse.com/wiki/Cosmological\_natural\_selection\_(fecund\_universes).

Freitas, R. 2000. *Some Limits to Global Ecophagy by Biovorous Nanoreplicators, with Public Policy Recommendations*. Foresight Institute Technical Report.

Goldblatt, Colin, and Andrew J. Watson. 2012. “The Runaway Greenhouse: Implications for Future Climate Change, Geoengineering and Planetary Atmospheres.” *Phil. Trans. R. Soc. A* 370 (1974): 4197–4216. doi:10.1098/rsta.2012.0004.

Grace, K. 2010a. “SIA Doomsday: The Filter Is Ahead | Meteuphoric.” *Meteuphoric*. https://meteuphoric.com/2010/03/23/sia-doomsday-the-filter-is-ahead/.

Grace, K. 2010b. “SIA Says AI Is No Big Threat.” *Meteuphoric*. https://meteuphoric.com/2010/11/11/sia-says-ai-is-no-big%c2%a0threat/.

Hanson, Robin. 1998. “Burning the Cosmic Commons: Evolutionary Strategies for Interstellar Colonization.” https://philpapers.org/rec/HANBTC.

Hippke, Michael, and John G. Learned. 2018. “Interstellar Communication. IX. Message Decontamination Is Impossible.” *ArXiv Preprint ArXiv:1802.02180*.

Hoffman, Paul F., Alan J. Kaufman, Galen P. Halverson, and Daniel P. Schrag. 1998. “A Neoproterozoic Snowball Earth.” *Science* 281 (5381): 1342–1346.

Holcombe, Larry, and Stanton T. Friedman. 2015. *The Presidents and UFOs: A Secret History from FDR to Obama*. New York: St. Martin’s Press.

Hoyle, Fred, John Elliot, and Alfredo Crespo. 1963. *A for Andromeda*. Plaza & Janes.

Inan, Umran S., Nikolai G. Lehtinen, R. C. Moore, K. Hurley, S. Boggs, D. M. Smith, and G. J. Fishman. 2007. “Massive Disturbance of the Daytime Lower Ionosphere by the Giant γ-Ray Flare from Magnetar SGR 1806–20.” *Geophysical Research Letters* 34 (8).

Jebari, Karim, and Niklas Olsson-Yaouzis. 2018. “A Game of Stars: Active SETI, Radical Translation and the Hobbesian Trap.” *Futures*.

Kahn, H. 1959. *On Thermonuclear War*. Princeton University Press.

Kardashev, Nikolai S. 1985. “On the Inevitability and the Possible Structures of Supercivilizations.” In , 112:497–504.

Kent, Adrian. 2004. “A critical look at risk assessments for global catastrophes.” *Risk Analysis* 24 (1): 157–68.

Korhonen, Janne M. 2013. “MAD with Aliens? Interstellar Deterrence and Its Implications.” *Acta Astronautica* 86: 201–210.

Lem, Stanis\law. 1999. *A Perfect Vacuum*. Northwestern University Press.

Lingam, Manasvi, and Abraham Loeb. 2017. “Risks for Life on Habitable Planets from Superflares of Their Host Stars.” *The Astrophysical Journal* 848 (1): 41.

Merali, Zeeya. 2016. *Shooting for a Star*. American Association for the Advancement of Science.

Mizokami, K. 2017. “That Time the U.S. Navy Had a Close Encounter With a UFO.” *Popular Mechanics*. https://www.popularmechanics.com/military/a14456936/that-time-the-us-navy-had-a-close-encounter-with-a-ufo/.

Napier, W. 2008. *Hazards from Comets and Asteroids. Global Catastrophic Risks*. Edited by N. Bostrom and Milan M. Cirkovic. Oxford University Press.

Panov, Alexander D. 2015. “Post-Singular Evolution and Post-Singular Civilizations.” In *Globalistics and Globalization Studies*, 361–376.

Reagan, R. 1987. “Address to United Nations General Assembly, 21 September 1987.” http://www.reagan.utexas.edu/archives/speeches/1987/092187b.htm.

Rhodes, Robert. 2014. “On The Potential Hostility of Alien Species.”

Saberhagen, Fred. 1984. *Berserker*. Ace Books.

Sandberg, A., S. Armstrong, and Milan M Cirkovic. 2017. “That Is Not Dead Which Can Eternal Lie: The Aestivation Hypothesis for Resolving Fermi’s Paradox.” *ArXiv Preprint ArXiv:1705.03394*.

Sandberg, A., Eric Drexler, and Toby Ord. 2017. “Dissolving the Fermi Paradox.” Future of Humanity Institute. http://www.jodrellbank.manchester.ac.uk/media/eps/jodrell-bank-centre-for-astrophysics/news-and-events/2017/uksrn-slides/Anders-Sandberg---Dissolving-Fermi-Paradox-UKSRN.pdf.

Sandberg, Anders, and Stuart Armstrong. 2013. “Hunters in the Dark: Game Theory Analysis of the Deadly Probes Scenario.” In *Poster Presented at the National Astronomy Meeting of the Royal Astonomical Society (NAM2013)*.

Smart, John. 2012. “The Transcension Hypothesis: Sufficiently Advanced Civilizations Invariably Leave Our Universe, and Implications for METI and SETI.” *Acta Astronautica Volume 78, September–October 2012, Pages 55–68*. http://www.sciencedirect.com/science/article/pii/S0094576511003304.

Smith, P.D. 2007. *Doomsday Men: The Real Dr. Strangelove and the Dream of the Superweapon*. St. Martin’s Press;

Smolin, Lee. 1992. “Did the Universe Evolve?” *Classical and Quantum Gravity* 9 (1): 173.

Sotala, Kaj, and Lukas Gloor. 2017. “Superintelligence as a Cause or Cure for Risks of Astronomical Suffering.” *Informatica* 41 (4).

Spry, J. Andy, Margaret Race, Gerhard Kminek, Bette Siegel, and Cassie Conley. 2018. “Planetary Protection Knowledge Gaps for Future Mars Human Missions: Stepwise Progress in Identifying and Integrating Science and Technology Needs.” In . 48th International Conference on Environmental Systems.

Srugatsky, N., and B. Strugatsky. 1985. *The Time Wanderers*. Richardson & Steirman.

Thrasymachus, K. 2012. “UFAI Cannot Be the Great Filter.” https://www.lesswrong.com/posts/BQ4KLnmB7tcAZLNfm/ufai-cannot-be-the-great-filter.

Tipler, Frank J. 1980. “Extraterrestrial Intelligent Beings Do Not Exist.” *Quarterly Journal of the Royal Astronomical Society* 21: 267–281.

Turchin, A. 2010. “Why Antropic Principle Stops to Defend Us Observation Selection, Future Rate of Natural Disasters and Fragility of Our Environment.” https://www.scribd.com/document/8729933/Why-anthropic-principle-stopped-to-defend-us-Observation-selection-and-fragility-of-our-environment.

Turchin, A. 2013. *UFO as Global Risk*.

Turchin, A. 2017. “Messaging Future AI.” *Manuscript*. https://goo.gl/YArqki.

Turchin, A. 2018. “The Risks Connected with Possibility of Finding Alien AI Code During SETI.” *Under Review in the Journal of the British Interplanetary Society*. https://www.academia.edu/30029491/The\_Risks\_Connected\_with\_Possibility\_of\_Finding\_Alien\_AI\_Code\_During\_SETI.

Turchin, A., and M. Chernyakov. 2018. “Classification of Approaches to Technological Resurrection.” *Under Review in Post-Humans Studies*. https://philpapers.org/rec/TURCOA-3.

Turchin, A., and D. Denkenberger. 2018a. “Classification of Global Catastrophic Risks Connected with Artificial Intelligence.” Under review in AI&Society.

Turchin, A., and D. Denkenberger. 2018b. “Surviving Global Risks through the Preservation of Humanity’s Data on the Moon.” *Acta Astronautica*, no. Volume 146, May 2018,: Pages 161-170.

Turchin, A., and D. Denkenberger. 2018c. “Global Catastrophic and Existential Risks Scale.” *Futures, in Press*.

Underdown, Simon. 2008. “A Potential Role for Transmissible Spongiform Encephalopathies in Neanderthal Extinction.” *Medical Hypotheses* 71 (1): 4–7.

Vallee, Jacques. 1969. *Passport to Magonia: From Folklore to Flying Saucers*. H. Regnery Company.

Vallee, Jacques. 1991. “Toward a Second-Degree Extraterrestrial Theory of UFOs: A Response to Dr. Wood and Prof. Bozhich.” *Journal of Scientific Exploration* 5 (1): 113–120.

Vinge, Vernor. 2013. *A Fire upon the Deep*. Hachette UK.

Ward, Peter D., and Donald Brownlee. 2003. *Rare Earth: Why Complex Life Is Uncommon in the Universe*. New York: Copernicus.

Wells, Herbert George. 1898. “The War of the Worlds.” *Random House Value Publishing. The Complete Science Fiction Treasury of HG Wells: 2003*.

Zaitsev, Alexander L. 2008. “Detection Probability of Terrestrial Radio Signals by a Hostile Super-Civilization.” *ArXiv Preprint ArXiv:0804.2754*.

# Appendix 1. Past civilizations on Earth

Based on known archaeological data, we are the first technological and symbol-using civilization on Earth (but not the first tool-using species). This leads to an analogy that fits the Fermi paradox: Why are we the first civilization on Earth? For example, flight was independently developed several times by evolution.

I will list possible explanations of the Fermi paradox of the human past and corresponding implications for existential risks (x-risks):

## **We are the first civilization on Earth, because we will prevent the existence of any future civilizations.**

The same idea was suggested by Berezin as explanation of the cosmic Fermi paradox (Berezin 2018).

If our existence prevents other civilizations from appearing in the future, how will that occur? We will either become extinct in a catastrophic way, killing all Earthly life, or become a super-civilization, which will prevent other species from becoming sapient. So, if we are really the first, it suggests that "mild extinctions" are not typical for human-style civilizations. Thus, pandemics, nuclear wars, devolutions and every type of reversible catastrophe is ruled out as a probable method of human extinction.

If we become a super-civilization, we may not be interested in preserving biospheres, as we will be able to create new sapient species. Or, it may be that we care about biospheres so strongly that we will be careful to hide from new sapient species as they appear, acting like the keepers of a cosmic zoo. Similarly, previous civilizations on Earth may have existed but decided to hide all traces of their existence from us to help us to develop independently.

The mere fact that we are the first raises the probability of a very large-scale catastrophe in the future, like UFAI, or dangerous physical experiments, and reduces the probability of mild x-risks, such as pandemics or nuclear war.

Another possible explanation is that any first civilization exhausts all of the resources that are needed for a technological civilization to start, such as oil, ores etc., so a second attempt is impossible. But over the course of tens of millions, years most such resources will be replenished or replaced.

## **2. We are not the first civilization**

2.1. We have not found any traces of a previous technological civilization, yet based on what we know, there are very strong limitations on their existence. For example, every civilization makes genetic marks, because it moves animals from one continent to another, just as humans brought dingoes to Australia. A civilization is also likely to exhaust several important ores, create artefacts, and create new isotopes. On this basis, we can be reasonably sure that we are the first technological civilization to exist on Earth in last 10 million years.

But, can we be sure about the past 100 million years? Maybe such a civilization was wiped out a very long time ago, like 65 million years ago (when a catastrophic event killed the dinosaurs). Carl Sagan argued that could not have happened, because we should find traces, mostly in the form of exhausted oil reserves. The main counterargument here cephalisation, the evolutionary development of the brains, was not advanced enough 60 million ago to support general intelligence. Dinosaurian brains were very small. However, birds’ brains are more mass effective than those of mammalians. All of these arguments are presented in detail in an excellent article by Brian Trent, “[Was there ever a dinosaurian civilization](http://www.strangehorizons.com/2009/20090713/trent-a.shtml)”?

The main x-risks here are that we will find dangerous artefacts from a previous civilization, such as weapons, nanobots, viruses, or AIs. And, if previous civilizations went extinct, that increases the probability that it is typical for civilizations to become extinct. It also means that there was some reason why an extinction occurred, and this killing force may be still active, and we might excavate it. If this civilization existed recently, its members were probably hominids (as some hominds were very advanced tool making species, like Homo Erectus), and thus, if they were killed by a virus, it may also affect humans.

2.2. We killed them. The Mayan civilization created writing independently, but Spaniards destroyed their civilization. Similar fates befell the Neanderthals and *Homo florentines*.

2.3. Myths about gods may be signs of such previous civilizations, but it is highly improbable, as such civilizations, even if they existed, extinct long before appearing of human cultural tradition.

2.4. They are still here, but they try not to intervene in human history in a manner similar to Fermi’s Zoo.

2.5. They were a non-technological civilization, and that is why we can’t find their remnants.

2.6 They may be still here, but like dolphins and ants their intelligence is non-human and they don’t create tech.

2.7 Some groups of humans created advanced technology long ago but prefer to hide it. Highly improbable, as most tech requires large manufacturing and market.

2.8 A previous humanoid civilization was killed by a virus or prion, and our archaeological research could bring it back to life. One hypothesis for the extinction of Neanderthals is prionic infection due to cannibalism (Underdown 2008). The fact is that several hominid species went extinct in the last several million years.

## **3. Civilisations are rare**

Millions of species have existed on Earth, but only one has created technology. So, it is a rare event. Consequences: cyclic civilizations on earth are improbable, so the chances that we will be resurrected by another civilization that arises on Earth is small.

The chances that we will be able to reconstruct civilization after a large-scale catastrophe are also small (as such catastrophes are atypical for civilizations and would quickly proceed to total annihilation or singularity).

It also means that technological intelligence is a difficult step in the evolutionary process, so it could be one of the solutions of the main Fermi paradox.

The safety of the remains of previous civilizations (if any exist) depends on two things: the distance in time from them and their level of intelligence. The greater the distance, the safer the remains (as the biggest part of dangerous technology will be destroyed by time, or will not be dangerous to humans, like species-specific viruses). The higher the level of the aliens’ intelligence, the riskier the remains they leave behind. *If anything appearing to be the remnants of an alien civilization is found, extreme caution is recommended*.

For example, the most dangerous scenario for us will be one similar to that described at the beginning Vinge’s novel “A Fire Upon the Deep.” We could find remnants of a very old, but very sophisticated civilization, which will include unfriendly AI or its description, or hostile nanobots. The most likely place for such artefacts to be preserved is on the Moon, in some cavities near the pole, as they are the most stable and radiation-shielded locations near the Earth.

I think that based on (no) evidence, an estimation of the probability of past technological civilizations should be less than 1 percent. While that is small enough to think that they most likely don’t exist, it is not enough to completely ignore the risk of their artefacts, though it is calculated as less than 0.1 percent.