Global Catastrophic Risks Connected with

Extra-Terrestrial Intelligence

August 2018

*Alexey Turchin*

Digital Immortality Now,

Foundation Science for Life Extension,

Corresponding author

alexeiturchin@gmail.com

*David Denkenberger*

University of Alaska Fairbanks

Alliance to Feed the Earth in Disasters (ALLFED)

david.denkenberger@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 extra-terrestrial intelligences (ETIs) 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 probability estimate, or that the Great Filter is ahead of us, and thus a future global catastrophe is high probability. If ETIs exist in our light cone, it is most likely they exist in the form of ET artificial general intelligence (ETAI). ETAI may be traveling at near-light speed as an explosive wave of intelligence in the form of some nano-mechanisms 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. This dormant form could be nanobots, which could act as berserkers, and be triggered when humans unwittingly cross some threshold. If ETIs are nearby and less advanced, they could find our messaging extra-terrestrial intelligence (METI) messages and send their “fleet” to or direct weapons toward Earth. Even extinct ETIs could have left dangerous remnants in the form of AI systems that our civilization may encounter. If our existence is a simulation, this could be run by ETIs. Even a false belief in the existence of ETIs could potentially fuel millennial sects. If humanity is the first intelligence to emerge, it may kill or prevent the existence of all future potential ETI, a different type of 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 ETIs will find the remains of our extinct civilization and resurrect it.

**Keywords**: extraterrestrial intelligence – SETI – METI – von Neumann probes – Fermi paradox – global catastrophic risks

**Highlights**:

* This paper 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 ETIs.
* The risks which are most often discussed are minuscule: sending messages attracting ETI and “alien invasion”.
* Humanity may be a threat to the emergence of future ETI.

# 1 Introduction

Stephen Hawking famously said that extraterrestrial intelligence (ETI) is one of the main risks to human existence [1]. In this article, we attempt explore all feasible ways in which aliens could affect global catastrophic risk (GCR), which is defined as catastrophes which could have an impact on the long-term future of humanity [2]. Paradoxically, even if ETI does not exist, humans may be still be in danger, as this suggests an increased probability of natural catastrophes like gamma-ray impacts and global warming.

The problems of the catastrophic risks of ETI are mostly explored in the scientific literature related to the problem of active search for extraterrestrial intelligence (SETI), or messaging extraterrestrial intelligence (METI), which could potentially attract the attention of aliens to us [3–5]. However, even passive SETI could be dangerous [6–8]. Korhonen has analyzed possible interstellar war scenarios and found them to be unlikely [9].

A major overview of the risks and benefits of the contact with ETI was carried out by Baum et al. [10]. Here, we present a different classification of the risks and include some new, related ideas, including an update of risk estimation based on an apparent lack of ETI, and the scenarios of alien simulation and non-sentient alien nanobots.

Another area of research is the Fermi paradox and the question of the location of a hypothetical Great Filter (Sandberg, Drexler, and Ord 2017), as many suggested types of future Great Filters are global catastrophes, e.g. geoengineering gone awry [12]. Explanations for the Fermi paradox may be divided into three broad groups: a) humans are alone in the observable universe; b) humans simply have not observed ETI yet; с) ETI is already here. Assuming that this list is exhaustive (c) would include the simulation hypothesis [Bostrom]), the probability of all three explanations combined is unity.

This article is 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 related to the existence of extraterrestrial intelligence as “ETI-risks”.*

In this article, we 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-risks. We 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 [8,9,13]. 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 the nature of technological progress and solutions of the Fermi paradox. We assume that advanced artificial intelligence (AI) with superhuman capabilities is possible based on the work of Bostrom [14], and that such AI are the natural outcome of the development of any civilization which does not self-destruct. We also assume that self-replicating nanorobots are also possible; and that travel near the speed of light based on these two technologies is possible [15]. We 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 [11].

ETI means 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 we will examine the idea that the non-existence of ETI means that something is killing off civilizations, and whatever is doing so may affect humanity as well.

## 2.1. Great Filter is behind us

The Fermi paradox is typically explained via the “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 [16] 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 humanity is lucky enough to live in a place where some types of catastrophe did not happen, including very large asteroid impacts [17], sterilizing gamma ray bursts [18], magnetar eruptions [19], very large solar flares [20], a runaway greenhouse effect [21], or a global snowball effect [22]. 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 much higher may not be 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 an 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 [24]. 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 ETI would decrease all the factors in the Drake equation, because N is now only 1. As we do not know which factor is the most significant, we should increase our probability estimate of all possible causes of the Early filter, 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 one 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 ET AI

K. 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 [25]. This conclusion requires some assumptions about how one calculates probabilities, that is, the self-indication assumption (SIA). Or one 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 [26], 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 [15].

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 nearly *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 [28]. 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—we have explained elsewhere why this could happen [29]. Unfriendly AI alone cannot be the explanation of the Fermi paradox, as it could start its own wave of space colonization via vPNs [30,31].

## 2.3. 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 [32]. 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 appears that our universe is fine-tuned for the creation of the intelligent life, and if we assume the logic of the fecund univers explanation, this should not be a coincidence. There are several theories of so-called cosmological natural selection with intelligence (CNS-I) [33]. In brief, it suggests that the interaction of a civilization with black holes is somehow necessary for universe to be replicated [34]. 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 with a particle accelerator of a small black hole [36]. Many such black holes could be created before they start to accrete earth mass in any measurable way, and each black hole may have a fecund universe inside it, which will explain the multiplication of the universes.

As space is silent, perhaps all civilizations destroyed themselves (but not the Universe) 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 probability 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 continued existence.

# 3. ETI exist in our light cone

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

Tipler [26] and later Armstrong et al. [15] 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. Such near-light vNPs may include nanorobots accelerated by something like the Breakthrough project’s solar sails [37] 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, ETAI 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 humans cannot understand will be discussed in section 7). As humanity still exists, the latter did not happen; thus, the wave has not reached us yet or is silent. As the wave is moving at near-light speed, humanity could not observe it much before it comes, and when it comes, it will likely 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 coauthors in the case of natural catastrophes like false vacuum decay [23]; 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 generally holds for the arrival of an intelligence explosion wave since many stars formed long before the sun.

However, there are several reasons that the wave is traveling significantly below the speed of light (e.g. because of the excessive energy requirements at near-light speed, interstellar medium friction or damage, the need of time to accelerate and decelerate probes as well as their self-replication, and the need to overcome the acceleration of the universe). For instance, if the speed were 0.9*c*, the wave may be observed long 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 humanity creates its own wave of AI capable of conquering a big part of the Galaxy, humanity may be safe from ET AI waves, as our wave will burn the cosmic commons necessary for an alien wave’s replication [38], like raw materials in planets; but if we exploited the energy potential of star systems over a long period of time, they could still be grabbed by a superior ETAI. 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: ET AI is downloaded via radio channel

If a remote ET 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 [39], as well as in an article by Carrigan [6]. [8] suggested 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 ET AI will need to be installed on Earth computers with human help as it could not directly infect the Internet, but obtaining such help will not be difficult for it, if it can be downloaded many times by many people.

ET AI could convert the Earth into another sending outpost, thus replicating and propagating further through the universe. Such messages should dominate between all SETI messages because they are replicating. As we develop more powerful radio telescopes and related instruments, we raise our chances of finding messages from ET AI, but we are still naïve, as we do not have our own general 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 advanced AI.

## 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. This seems improbable, because if they are so near, why wouldn’t they have reached Earth yet given that many stars are much older than the sun? This only plausible if their technology is not very advanced and cannot evolve.

Korhonen has explored the idea of such a preemptive strike by ETI [9]. However, it appears 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, Earth is either inside or outside a sphere created by the wave. If we are inside, ETI is likely to be present in the Solar system. If we outside, it is very unlikely that the wave hit us before we reach Kardashev-2 level. For example, if the median distance between civilizations is a billion light years, and the wave has 0.5c speed and the time until our Kardashev level is 1000 years, the chance that the wave hit us during our period of vulnerability is 2x10-6.

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 [9]. 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 [40,41].

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 hundred light years or so 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 [9]. 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 ET zombies

It is possible that ETs 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. One risk we have not seen discussed is if ET created self-replicating non-sentient nanobots (grey goo), they could travel as interstellar stardust and convert all solid matter they encounter to nanobots, so Earth 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 humanity. If the ET 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 [31]. Their damaged rationality could make them extremely single-minded in pursuing absurd goals.

If humanity finds 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” [42].

ETI may have visited the Solar system hundreds of millions of years ago. One of the places where their remains may be found is the Moon, which is very geologically stable [43]. Ideas about possible space archeology have already been suggested [44,45].

ETs 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 ET bacteria resulting in a deadly pandemic [10]. We find such a scenario unlikely, as visits by large starships seem less probable than visits by small, self-replicating probes, and because ET biology may be wholly incompatible with ours.

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

By messaging to ET (METI) in order to initiate communication, humanity could tell potentially hostile ETs our position in space. Some people advocate for METI, including Zaitsev [3], while others are strongly opposed, including Brin [4]. The risks of METI are smaller than risks of SETI, in our opinion, as human radio signals can only reach the nearest few hundred light years until humanity will create our own superintelligent AI which will be able to provide advance space defense.

Incoming messages found via SETI could come from a much greater distance, where the existence of more advance civilization is more probable. In the case of SETI, there is no time delay for the attack, as the sending of alien signals could have started even before our species appeared. The ETI would send the message in the hope that someone would eventually encounter the contaminated message, as humanity could find such signals as soon as it has the needed observational instruments. In the case of METI, there will be a delay, because the signal has to travel 2 ways: first, from Earth to the ETI, and then back. For example, if the ETIs is 100 ly from us and they were capable to detect our fist radio transmissions around 1900, the earliest time of their observable reaction is 2100.

In the case of SETI, humanity is able to receive signals from much greater distances, perhaps as far as one billion light years. So, SETI program is exposed to a much larger part of the Universe (perhaps as much as 1010 times the number of stars) than humanity could reach via METI. SETI-attack is not a “physical attack” but just a contaminated informational signal. METI-induced-attack may be physical, but requires much higher density of ETIs. METI also requires strange passivity in space exploration, as any ETI capable of attacking Earth physically would likely be able to visit our Solar system long before now.

Most nearby ETs, if any exist, will at least know about the habitability of our world via exoplanet observation in telescopes. Thus, they could guess existence of life and then 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 ETs which are at the same level of development. Such encounters may result in classical star wars. [9] 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 [46] may include weapons of enormous power, including several which one 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, and
* 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 which lay dormant and might never got triggered, like land mines; or some form of dangerous replicators, like space “grey goo”.

## 3.7. Superintelligent civilizations acting on our world from a great distance

As S. 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 [47]. 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. ETs prefer to be silent in order not to attract the attention of some unknown force (maybe UFAI) or because of fear of mutual hostilities [9]. If this is the case, humanity should also be silent until we will learn what is going on.

# 4. ETs are here

## 4.1. ET 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.

ET 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 [48,49]. 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 [50]. Such activity may be explained by the activity of ET 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 ETs as berserkers

If Earth is located inside the sphere of the intelligence explosion wave of an ET AI, it 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 ET nanotech will have to prevent the creation of self-improving AI by humanity.

If there are ETs 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”.

ET berserkers could take the form of remote probes observing planets from space for signs of intelligent life, like radioemissions. Given the possibility of ET nanotechnology, they could be not very far away, as they could be very small and well disguised as natural objects, like small asteroids.

## 4.3. “Bona fide” ET invasion is very unlikely

It has been a popular trope since Wells’ “The War of the Worlds” [54] that non-AI ETs 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 ET arrival is likely fairly 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 ETs will travel between stars as biological bodies. Finally, 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 [26]. However, ET may have non-human value system which make difficult to predict their behavior rationally.

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

It seems that the taboo of scientific discussion of the unexplainable phenomena in the sky is slightly relaxing with several scientific and mainstream media publications [55,56].

However, it is unlikely that observed unidentified flying objects (UFOs) are ET spaceships because sending very large starships through vast cosmic distances is a very impractical idea [57]. 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” [58]. Each explanation of UAP tells us more about the epoch than about actual observations. Humanity could use our current epoch’s explanatory potential to rationalize UAP as glitches in the Matrix, clouds of ET 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 [59]:

1. **Hoaxes, hallucinations and optical illusions**. This is the most accepted explanation for UAPs. The greatest 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 [60]. False beliefs in ETs may result in incorrect decisions. He also mentioned ET 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” [61].
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 ET starships. The risks are possible war with whatever the entities are or the consequences of obtaining new dangerous technologies. But the biggest risks are likely unknown and unknowable.

## 4.5. We are in a simulation created by ETs, 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 ETs 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 more often (or one could say that ETs themselves are “players” and thus have some presence within the simulation). To get a good model of the late Great Filter, the ET 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. That is, unless there are no other dominating types of simulation or unless the proportion of naturally extinct civilizations to supercivilizations is very high (as there will be more extinct civilizations than simulations created by civilizations that continued)—but in that case, humanity faces 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 [63].

## 4.6. The “space zoo”

A “space zoo” is an area protected by ETI for ethical, scientific or recreational proposes. If humanity is in a space zoo, it actually increases our chances of survival, as ETI are likely to prevent some global catastrophes. But ET help for us may be completely foreign from our ethical point of view and thus unacceptable [64].

# 5. Prevention of GCRs with help of ETI

## 5.1.   Resurrection by ETI

We could preserve some information about humanity in the hope that ETs will resurrect us, or they could return us to life using our remains on Earth [43]. 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 [43].

## 5.2.   Request for help

Humanity could send radio messages with a request for help. (This would only be a gesture of desperation if they are not already hiding in the Solar system, because if they are on stellar distances, the reply will come at least in decades, when it will be – likely – late). People might find advice on how to prevent GCRs in ET messages received via SETI [10]. Panov suggested that SETI will help us to join an “galactic internet” of exo-humanistic civilizations [65].

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

# 6. We are the risk for ETs

There is also a risk that humans will cause the extinction of ETs or prevent their future appearance. If we assume unbounded altruism, which Baum et al. called universalist ethics [10], killing other sentient beings is bad. Also, if one assumes functional decision theory [66], our decision not to kill ETs may mean that ETs who use the same line of reasoning, will also avoid acting in ways that are dangerous to humans.

One way in which humanity could affect all other ET civilizations in the Universe is by starting a space colonization wave, which could be non-sentience self replicators or an intelligence explosion wave with our AI (whether friendly or not). Another way is via a failed accelerator experiment in which scientists create false vacuum decay, affecting all the universe in our light cone (an artificial black hole or stranglet would destroy only Earth [36]). This decay catastrophe would be worse than most existential risks from a moral point of view, as it would kill many more sentient beings [67].

Another risk to ETs is space colonization, where humanity does not directly kill the ETI but prevents their future appearance by consuming or colonizing potentially habitable planets. For example, if humans colonize Mars via terraforming, they may kill any bacterial life that may exist under its surface, and such risks are already taken into consideration in space exploration [68].

Berezin [69] recently published an article with the main idea that the most obvious solution to the Fermi paradox is that humanity will kill all ETs, because only one civilization can exist in any light cone. Thus, humanity would prevent existence of all other civilizations via colonization or military self-protection measures.

# 7. Non-human intelligence

Everything set above is based on assumption of “ETI rationality”, that is that ET goals and the ways to achieve them are comprehensible to humans. However, this may be not true. There are several ways that ET irrationality could appear. This non-rationality may be especially important in explanation of different zoo scenarios and is in fact the narrative engine of many works of science fiction: e.g. Lem’s *Solaris* (Lem 1970).

## 7.1 ET goals are rational but incredibly complex

There are many types of human behavior which are incredibly complex and cannot be explained without context, like contemporary art. In the movie “Predator” ETs preserve the Earth as sanctuary, but only to go there for hunting.

## 7.2. ETs are mad

ET remains may survive as brain-damaged robots capable to self-replicating. A “paperclip maximizer” is an example of such craziness in an AI system, which has non-human aligned goals and only wants to create infinite number of paperclips [14].

## 7.3. ETs have rational goals but non-human way of achieving them

Natural selection of Darwinian evolution is an example of a non-human way of problem solving. While in the long term, it can create complex mechanisms, one cannot 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 ETs are generally inverse to the popularity of the relevant media tropes. The biggest risk is the non-existence of ETs, as this would mean a high probability for our own extinction. The next is pure informational and passive listening to ETs, as it makes us vulnerable to Trojan messages from an ET 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 ET 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”.

# References

[1] C. Cofield, Stephen Hawking: Intelligent Aliens Could Destroy Humanity, But Let’s Search Anyway, Space.Com. (2015). https://www.space.com/29999-stephen-hawking-intelligent-alien-life-danger.html (accessed February 17, 2018).

[2] N. Bostrom, M.M. Cirkovic, Global catastrophic risks, Oxford University Press, 2011.

[3] A.L. Zaitsev, Detection probability of terrestrial radio signals by a hostile super-civilization, ArXiv Prepr. ArXiv08042754. (2008).

[4] D. Brin, The dangers of first contact: the moral nature of extraterrestrial intelligence and a contrarian perspective on altruism, Skept. Altadena CA. 15 (2010) 28–36.

[5] K. Jebari, N. Olsson-Yaouzis, A Game of Stars: Active SETI, Radical Translation and the Hobbesian Trap, Futures. (2018).

[6] R.A. Carrigan Jr, Do potential SETI signals need to be decontaminated?, Acta Astronaut. 58 (2006) 112–117.

[7] M. Hippke, J.G. Learned, Interstellar communication. IX. Message decontamination is impossible, ArXiv Prepr. ArXiv180202180. (2018).

[8] A. Turchin, The Risks Connected with Possibility of Finding Alien AI Code During SETI, J. Br. Interplanet. Soc. 70 (2018).

[9] J.M. Korhonen, MAD with aliens? Interstellar deterrence and its implications, Acta Astronaut. 86 (2013) 201–210.

[10] S.D. Baum, J.D. Haqq-Misra, S.D. Domagal-Goldman, Would contact with extraterrestrials benefit or harm humanity? A scenario analysis, Acta Astronaut. 68 (2011) 2114–2129.

[11] A. Sandberg, E. Drexler, T. Ord, Dissolving the Fermi Paradox, (2017). 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 (accessed December 28, 2018).

[12] M.M. Cirkovic, R. Cathcart, Geo-engineering Gone Awry: A New Partial Solution of Fermi’s Paradox, ArXiv Prepr. Physics0308058. (2003).

[13] R. Rhodes, On The Potential Hostility of Alien Species, (2014).

[14] N. Bostrom, Superintelligence, Oxford University Press, Oxford, 2014.

[15] S. Armstrong, A. Sandberg, Eternity in six hours: intergalactic spreading of intelligent life and sharpening the Fermi paradox, Acta Astronaut. 89 (2013) 1–13.

[16] P.D. Ward, D. Brownlee, Rare Earth: Why Complex Life is Uncommon in the Universe, Copernicus, New York, 2003.

[17] W. Napier, Hazards from comets and asteroids. Global Catastrophic Risks, Oxford University Press, 2008.

[18] M.M. Ćirković, B. Vukotić, Long-term prospects: Mitigation of supernova and gamma-ray burst threat to intelligent beings, Acta Astronaut. 129 (2016) 438–446. doi:10.1016/j.actaastro.2016.10.005.

[19] U.S. Inan, N.G. Lehtinen, R.C. Moore, K. Hurley, S. Boggs, D.M. Smith, G.J. Fishman, Massive disturbance of the daytime lower ionosphere by the giant γ-ray flare from magnetar SGR 1806–20, Geophys. Res. Lett. 34 (2007).

[20] M. Lingam, A. Loeb, Risks for life on habitable planets from superflares of their host stars, Astrophys. J. 848 (2017) 41.

[21] C. Goldblatt, A.J. Watson, The runaway greenhouse: implications for future climate change, geoengineering and planetary atmospheres, Phil Trans R Soc A. 370 (2012) 4197–4216. doi:10.1098/rsta.2012.0004.

[22] P.F. Hoffman, A.J. Kaufman, G.P. Halverson, D.P. Schrag, A Neoproterozoic snowball earth, Science. 281 (1998) 1342–1346.

[23] M.M. Ćirković, A. Sandberg, N. Bostrom, Anthropic shadow: observation selection effects and human extinction risks, Risk Anal. Vol 30 No 10 2010. (2010).

[24] A. Turchin, Why antropic principle stops to defend us Observation selection, future rate of natural disasters and fragility of our environment, (2010). https://www.scribd.com/document/8729933/Why-anthropic-principle-stopped-to-defend-us-Observation-selection-and-fragility-of-our-environment (accessed December 28, 2018).

[25] K. Grace, SIA doomsday: The filter is ahead | Meteuphoric, Meteuphoric. (2010). https://meteuphoric.com/2010/03/23/sia-doomsday-the-filter-is-ahead/ (accessed August 10, 2018).

[26] F.J. Tipler, Extraterrestrial intelligent beings do not exist, Q. J. R. Astron. Soc. 21 (1980) 267–281.

[27] N. Bostrom, Existential risks: Analyzing human extinction scenarios and related hazards, (2002).

[28] N. Bostrom, The Vulnerable World Hypothesis, (2018) 38.

[29] A. Turchin, D. Denkenberger, Classification of Global Catastrophic Risks Connected with Artificial intelligence, (2018).

[30] K. Thrasymachus, UFAI cannot be the Great Filter, (2012). https://www.lesswrong.com/posts/BQ4KLnmB7tcAZLNfm/ufai-cannot-be-the-great-filter (accessed August 10, 2018).

[31] K. Grace, SIA says AI is no big threat, Meteuphoric. (2010). https://meteuphoric.com/2010/11/11/sia-says-ai-is-no-big%c2%a0threat/ (accessed August 10, 2018).

[32] L. Smolin, Did the universe evolve?, Class. Quantum Gravity. 9 (1992) 173.

[33] Evo-devo wiki, Cosmological natural selection (fecund universes), (2018). http://evodevouniverse.com/wiki/Cosmological\_natural\_selection\_(fecund\_universes) (accessed February 8, 2018).

[34] L. Crane, Possible implications of the quantum theory of gravity: An introduction to the meduso-anthropic principle, Found. Sci. 15 (2010) 369–373.

[35] J. Smart, The transcension hypothesis: Sufficiently advanced civilizations invariably leave our universe, and implications for METI and SETI, Acta Astronaut. Vol. 78 Sept. 2012 Pages 55–68. (2012). http://www.sciencedirect.com/science/article/pii/S0094576511003304.

[36] A. Kent, A critical look at risk assessments for global catastrophes, Risk Anal. 24 (2004) 157–68.

[37] Z. Merali, Shooting for a star, American Association for the Advancement of Science, 2016.

[38] R. Hanson, Burning the cosmic commons: evolutionary strategies for interstellar colonization, (1998). https://philpapers.org/rec/HANBTC (accessed December 28, 2018).

[39] F. Hoyle, J. Elliot, A. Crespo, A for Andromeda, Plaza & Janes, 1963.

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

[41] H. Kahn, On thermonuclear war, Princeton University Press, 1959.

[42] V. Vinge, A fire upon the deep, Hachette UK, 2013.

[43] A. Turchin, D. Denkenberger, Surviving global risks through the preservation of humanity’s data on the Moon, Acta Astronaut. (2018) Pages 161-170.

[44] R.A. Carrigan Jr, Is interstellar archeology possible?, Acta Astronaut. 78 (2012) 121–126.

[45] P.C.W. Davies, R.V. Wagner, Searching for alien artifacts on the moon, Acta Astronaut. 89 (2013) 261–265.

[46] N.S. Kardashev, On the inevitability and the possible structures of supercivilizations, in: Serch Extraterr. Life, 1985: pp. 497–504.

[47] S. Lem, A perfect vacuum, Northwestern University Press, 1999.

[48] Drexler, E.: Engines of Creation, Anchor Press, 1986.

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

[50] J.E. Enriquez, A. Siemion, T.J.W. Lazio, M. Lebofsky, D.H. MacMahon, R.S. Park, S. Croft, D. DeBoer, N. Gizani, V. Gajjar, Breakthrough Listen Observations of 1I/′ Oumuamua with the GBT, Res. Notes AAS. 2 (2018) 9.

[51] A. Sandberg, S. Armstrong, M.M. Cirkovic, That is not dead which can eternal lie: the aestivation hypothesis for resolving Fermi’s paradox, ArXiv Prepr. ArXiv170503394. (2017).

[52] F. Saberhagen, Berserker, Ace Books, 1984.

[53] A. Sandberg, S. Armstrong, Hunters in the dark: Game theory analysis of the deadly probes scenario, in: Poster Present. Natl. Astron. Meet. R. Astonomical Soc. NAM2013, 2013.

[54] H.G. Wells, The War of the Worlds., Random House Value Publ. Complete Sci. Fict. Treas. HG Wells 2003. (1898).

[55] A. Dodd, Strategic Ignorance and the Search for Extraterrestrial Intelligence: Critiquing the Discursive Segregation of UFOs from Scientific Inquiry, Astropolitics. 16 (2018) 75–95. doi:10.1080/14777622.2018.1433409.

[56] H. Cooper, R. Blumenthal, L. Kean, Glowing Auras and ‘Black Money’: The Pentagon’s Mysterious U.F.O. Program, N. Y. Times. (2017). https://www.nytimes.com/2017/12/16/us/politics/pentagon-program-ufo-harry-reid.html (accessed August 28, 2018).

[57] J. Vallee, Toward a second-degree extraterrestrial theory of UFOs: a response to Dr. Wood and Prof. Bozhich, J. Sci. Explor. 5 (1991) 113–120.

[58] J. Vallee, Passport to Magonia: From folklore to flying saucers, H. Regnery Company, 1969.

[59] A. Turchin, UFO as Global Risk, 2013.

[60] K. Mizokami, That Time the U.S. Navy Had a Close Encounter With a UFO, Pop. Mech. (2017). https://www.popularmechanics.com/military/a14456936/that-time-the-us-navy-had-a-close-encounter-with-a-ufo/ (accessed August 11, 2018).

[61] R. Reagan, Address to United Nations General Assembly, 21 September 1987, (1987). http://www.reagan.utexas.edu/archives/speeches/1987/092187b.htm.

[62] N. Bostrom, Are You Living In a Computer Simulation?, Publ. Philos. Q. 2003 Vol 53 No 211 Pp 243-255. (2003).

[63] A. Turchin, M. Chernyakov, Classification of Approaches to Technological Resurrection, Rev. Post-Hum. Stud. (2018). https://philpapers.org/rec/TURCOA-3 (accessed December 28, 2018).

[64] N. Srugatsky, B. Strugatsky, The Time Wanderers, Richardson & Steirman, 1985.

[65] A.D. Panov, Post-singular evolution and post-singular civilizations, in: Glob. Glob. Stud., 2015: pp. 361–376.

[66] E. Yudkowsky, N. Soares, Functional Decision Theory: A New Theory of Instrumental Rationality, ArXiv171005060 Cs. (2017). http://arxiv.org/abs/1710.05060 (accessed August 18, 2018).

[67] A. Turchin, D. Denkenberger, Global Catastrophic and Existential Risks Scale, Futur. Press. (2018).

[68] J.A. Spry, M. Race, G. Kminek, B. Siegel, C. Conley, 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, 2018.

[69] A. Berezin, “First in, last out” solution to the Fermi Paradox, ArXiv180308425 Phys. (2018). http://arxiv.org/abs/1803.08425 (accessed May 29, 2018).

[70] S. Lem, Solaris. 1961, Trans Steve Cox Joanna Kilmartin N. Y. Hartcourt Brace. (1970).