Journal of Evolution & Technology

A peer-reviewed electronic journal published by the Institute for Ethics and Emerging Technologies

ISSN 1541-0099

25(1) – June 2015

Extraterrestrial artificial intelligences and humanity's cosmic future: Answering the Fermi paradox through the construction of a Bracewell-Von Neumann AGI

Tomislav Miletić
tomislav.miletich@gmail.com

Journal of Evolution and Technology - Vol. 25 Issue 1 – June 2015 - pgs 56-73

Abstract

A probable solution of the Fermi paradox, and a necessary step in humanity's cosmic development, is the construction of a Bracewell-Von Neumann (BN) Artificial General Intelligence (AGI). The use of BN probes is the most plausible method of initial galactic exploration and communication for advanced ET civilizations, and our own cosmic evolution lies firmly in the utilization of, and cooperation with, AGI agents. To establish these claims, I explore the most credible developmental path from carbon-based life forms to planetary civilizations and AI creation. I consider the likely physical characteristics of extraterrestrial AI probes and propose ways to predict their behavior. Lastly, I ponder the possible trajectories for humanity's cosmic future.

1. Introduction: the Fermi paradox

When Fermi famously stated his question in the 1950s, advanced extraterrestrials were imagined as beings of flesh and blood, exploring the galaxy in saucer-like space ships, and beaming signals from their home worlds. A decade later, the SETI project was initiated with the hope of receiving a transmission from a distant ET civilization through radio communications (Cocconi and Morrison 1959). Unfortunately, the radio silence (Brin 1983) persists. Without a decidedly unambiguous signal of intelligent ET origin, we are forced to consider other approaches to SETI, especially those that are in line with contemporary advances in sciences. These approaches can be separated into two groups: planetary and extraplanetary (re)search.

The planetary approach includes such novel and noteworthy examples as inspecting our DNA for encoded messages, searching for a shadow alien biosphere, and listening for ET on the internet (Harrison 2010). The extraplanetary approach focuses on the search for a general class of technological artifacts as well as manifestations and products of advanced ET civilizations in our planetary vicinity and beyond (Bradbury et al. 2011). Although the planetary approach remains valuable, our best chances may lie with the extraplanetary approach. Within that context, it is time for us to move on from the radio paradigm. As numerous voices have suggested, the most efficient means of intergalactic communication would most likely occur by means of a kind of space-faring artificial intelligence usually termed the Bracewell-Von Neumann probe. Since this is one of the most likely galactic scenarios, we should commence a collaboration between the AI and SETI fields to ensure the possibility of creating our own Bracewell-Von Neumann agent in the foreseeable future.

In the following pages, I will first take a brief look at the current state of astronomical sciences. The aim here is to establish the galactic requirements and most probable outcomes of planetary biological evolution, especially with regard to intelligent life. Second, I will identify the requirements for a technological civilization and argue that the driving factor of cultural evolution – the "Intelligence Principle" – should guide every ET civilization toward the goals of space exploration, AI creation, and possibly even postbiology. Before concluding this discussion, I will examine the various possibilities of ET culture that should be reflected, at least in part, in the programming of extraterrestrial artificial intelligence (ETAI). Lastly, I will ponder the goals of our own AI research within the context of galactic exploration and ET communication.

2. From biological to artificial intelligence – the probable path

2.1. Biophysical requirements

The past decades have shown tremendous progress in astrobiology, which aims to answer the questions of "how life originated on Earth, whether there is life elsewhere in our solar system and beyond, and what the future holds for life" (O'Malley-James and Lutz 2013, 95). We have come to learn that, for life to develop, certain specific conditions are needed. Sciences such as geology, geochemistry, astronomy, and planetology have helped us establish the requirements for life's emergence and sustainment, and evolutionary sciences explain the possibilities of multicellular evolution. Although we have no other example with which to compare our planet's biological and cultural history, we are aware of some crucial points required for the evolution of life and the rise of an intelligent civilization capable of space travel. I will try to portray this narrative by showing that there exists a universal evolutionary route from carbon-based biochemistry to intelligent planetary life and its cultural evolution leading to extraplanetary and galactic exploration through the creation of artificial intelligences.

Let's start with the basics. Life needs to be able to carry, transform, and inherit information. For this, a physical element with a complex structure is required. As far as we know, carbon is the most suitable of the available elements.

There are numerous reasons for this. One is carbon's ability to form bonds with other atoms. The ability to interactively engage chemical bonds – and particularly to form double or triple bonds with other atoms – allows carbon to become highly present (more than 75 per cent) within the entirety of interstellar and circumstellar molecules (Henning and Salama 1998). This allows it to form an extraordinary range of complex molecules. For example, polycyclic aromatic hydrocarbons (PAHs) contain 2-10% of all carbon in space. Indeed, PAHs are among the most common and abundant polyatomic molecules in the visible universe (Ehrenfreund and Sephton 2006).

Significantly, when subjected to UV radiation, PAHs are transformed into biogenic complex organics, making them one of the best possible candidates for life's initial building blocks (Wickramasinghe and Trevors 2013). Additionally, simulations in prebiotic chemistry and meteoritic carbon monomer findings testify that carbon-based biochemical traits (homochirality, α amino acid configuration, β sheet structure) represent universal motifs of life (Davila and McKay 2014).

Still, there could exist alternative biochemistries such as those based on silicon, since silicon also has the ability to form sufficiently large macromolecules. Because it requires liquid ammonia or nitrogen as a solvent, some scientists even look toward Saturn's largest moon, Titan, theorizing a possible surface chemistry based on silicon (Bains 2004). But there are major obstacles to silicon-based life. For example, silicon lacks the ability to form double bonds with as wide an array of atoms as carbon, resulting in decreased distribution and complexity-forming capabilities. However, the most important obstacle for silicon-based life is that silicon cannot use water as a solvent.

The importance of water lies not only in the primary role it provides for the evolution of carbon-based organisms, but also in a crucial geological effect: the softening and deforming of the lithosphere, which results in the subduction of the crust and the formation of plate tectonics. Recent discoveries have revealed the importance of plate tectonics as one of the most important geological systems for the emergence and sustainability of complex life on planets (Fishbaugh et al. 2007).

Specifically, the carbonate-silicate cycle keeps the atmospheric volume of carbon dioxide relatively uniform, which in turn maintains the temperature range required for the appearance and evolution of multicellular life. Plate tectonics also keeps the magnetic field operational by cooling the planet's interior.

A magnetic field's importance lies in its ability to reroute dangerous extraplanetary radiation including solar wind exposure. If Earth's magnetic field were to stop working, the atmosphere would erode and allow UV radiation to punch through, heating up the mantle and destabilizing surface liquid water. Needless to say, all of this would reduce the chances for life's arrival. If life is already present on a planet, it would severely limit the diversification of the planet's biosphere.

In addition to the necessity of water and a stable temperature range (Rospars 2010), the evolution of carbon-based life requires an abundant and easily available energy source. The best source is a solar one. Although other sources, such as geothermal or radioactive disequilibria, could support multicellular evolution, they simply cannot match the output of solar energy (Benner et al. 2010). To use solar energy effectively, evolution employs the photosynthetic approach, which utilizes a quantum wavelike process for maximum efficiency in energy utilization (Engel et al. 2010). In ensuring the highest energy intake, photosynthesis supports a greater development of complexity and productivity, and greater biological diversity both on the seafloor and in water (O'Malley-James and Lutz 2013; Nisbet et al. 2007). The photosynthetic process also opens a superior path for multicellular development, because it creates most of the planetary oxygen as well as biologically useful carbon (Iverson 2006, 97), while also providing the energy required by multicellular organisms with various skeletal structures and modes of locomotion (Cockell 2007; Knoll 2003; Paine 2011). Ultimately, the development of photosynthesis may also be required for the evolution of intelligence.

2.2. Evolution of intelligence, culture, and technology

If life is to survive and develop into *intelligent* life, it needs to adapt to unstable environments and to generate additional complexity. For this to happen, it needs to evolve through the process of natural selection (Dawkins 2010, 371) or perhaps by a process of "environmental conditions that are continuously creating different life forms, or similar life forms with adapted traits" (National Research Council 2010, 212). Since the latter option should be taken as extremely rare, or even faulty, when compared to the mechanism of natural selection, natural selection is "the fundamental mechanism for the evolution of initial life forms and subsequently intelligence in the universe" (Bedau and Cleland, 120).

This leads us to conclude that purely informational or synthetic robotic life forms are not expected to arise from the evolutionary process and are not the starting point of intelligence. Rather, they are products of additional cultural and technological evolution. In other words, synthetic, robotic life forms are produced by carbon-based evolved intelligences and not the other way around.

Still, even if intelligence is most likely to be achieved through a process of carbon-based biological evolution, we cannot claim with certainty that intelligence is a convergent feature of universal evolution (Rospars 2010). What we can affirm is that intelligence, as a product of the evolutionary process, is a step in the advancement of biosystems that develop complex structures with greater diversity, energy, and hierarchical levels (Ekstig 2010; Toussaint and Schneider 1998; Tessera and Hoelzer 2013). Our own example shows that human-like intelligence is not entirely qualitatively different from the intelligence of other animals on our planet:

> Contributions in ethology and animal psychology have recognized aspects of imitation, theory of mind, grammatical–syntactical language and consciousness in non-human primates and other large-brained mammals as well. (Roth and Dicke 2005, 256)

The difference between human-like intelligence and animal intelligence lies in the level of evolved abilities, particularly the ability for temporal analysis with motor behavior, action planning, thinking, and language (Macphail and Bolhuis 2001; Fuster 2002). These evolutionary advantages allowed our species to find and extract energy from difficult-to-acquire high calorie foods, which fostered the growth in body and brain size, thereby fueling an additional rise of intelligence.

Likewise, biological improvements changed social life, with increased longevity, prolonged maturation, and a large commitment to learning inside a socially organized group (Kaplan et al. 2000). Consequently, social interactions gave rise to selection pressures for advanced cognition, "supporting the view that the transition to the cooperative groups seen in the most intelligent species on our planet may be the key to their intellect" (McNally et. al. 2012, 3033). Finally, social cooperation led to further improvement and transfer of technological knowledge to future generations (Bjorklund 2006).

Technology is therefore as much a cultural force as it is an evolutionary development. But for a technological civilization to arise, certain planetary conditions are needed. Luckily, these are the same as those required for the evolution of complex intelligent life. They include a planet with a metal core, an abundance of metals throughout the core, solid ground with the availability of solar power as the most efficient, abundant, and long lasting energy resource, and a stable climate.

It is plausible to reason that if the conditions on other planets are life-friendly, “life forms might evolve in hierarchical organization, size, diversity and information-processing skills” (Rospars 2013, 19). Additionally, if intelligence is common, we may as well be living in a postbiological universe “in which flesh-and-blood intelligence has been either augmented, replaced or substituted by artificial intelligence” (Dick 2009, 578). The reason for this opinion lies in the Intelligence Principle:

> the maintenance, improvement and perpetuation of knowledge and intelligence is the central driving force of cultural evolution, and that to the extent intelligence can be improved, it will be improved…. The Intelligence Principle implies that, given the opportunity to increase intelligence (and thereby knowledge), whether through biotechnology, genetic engineering or artificial intelligence, any society would do so, or fail to do so at its peril. (Dick 2009, 579)

But, are ET cultures poised to pursue a postbiological future? The answer is difficult to find, since we have no complete theories of even our own cultural evolution and its mechanisms. We also cannot predict whether an alien culture would reject the postbiological phase for religious or philosophical reasons.

It is safe, nonetheless, to claim that all ET cultures will pursue species survival through resource acquisition and growth in intelligence. Since planetary survival is constantly endangered by cosmic and planetary calamities, including species-induced ecological disasters, the survival instinct will propel every sentient species beyond the confines of its own planet toward extraplanetary colonization. Unfortunately, space conditions are detrimental and lethal to carbon-based lifeforms (Harrison 2010).

Thus, if a technological civilization is to maximize the odds of its survival through space exploration and planetary colonization, it will need to develop forms that can survive the effects of prolonged exposure to space environments. An intelligent thinking machine capable of space travel, communication, and tool use is the most probable of such options, and we can safely guess that a distant alien civilization would initially explore the galaxy through a certain kind of ETAI.

The most probable of such agents is the self-replicating “Bracewell-von Neumann” (BN) probe. The scenario for such a probe requires the oldest possible alien civilization, one that could have evolved several billion years ago in the Milky Way Galaxy (Dick 2009). When a civilization enters the technological phase required for galactic exploration, it will first survey the galaxy to find planets residing in habitable zones. Its next step is to count the number of those planets, calculate the distances between them, and proceed with dispatching BN probes. The task of an intelligent probe is to enter a designated solar system and initiate its programmed goals. Since it stays in the planet’s vicinity, it has no need for high energy consumption. The proximity of the probe shortens the communication to light-minutes while not revealing the home location of the probe’s sender. Upon arrival, the probe can passively monitor any local technological society before initiating contact. To remain functionally intact, the probe will need to have an intelligent ability for self-repair and the ability for self-manufacturing. Required materials and energy can be harvested from raw materials in space and the designated solar system.

But if BN machines are one of the most efficient agents (in terms of energy usage, building costs, and time consumption) of galactic communication, and if it is logical to assume that they would be widely used by ET civilizations, why haven’t we come into contact with one of them? One possible reason is, as always, that we are alone in our galaxy. Frank Tipler has claimed that the galaxy's colonization by these machines would take around 300 million years and that their absence from our solar system represents a more potent version of the Fermi paradox arguing against the existence of ETs (Davies 2010, 74).

Since we have only recently begun exploring our solar system, we cannot take the absence of BN probes as a matter of fact. In fact, just the opposite could be true – the BN could be well hidden in a “secret” location and waiting to reveal itself if we fulfill a certain expected condition (Gillon 2013). Or perhaps we need to search in the “right” direction or the “right” way to demonstrate that we have achieved a certain technological or cultural level. Or perhaps we need a different kind of mind to help us discover an alien mind.

It is in our best interests to mitigate the *unknown* factor as much as possible while we contemplate an ETAI agent’s possible existence. The “Titanic effect” occurs “when we are so certain that an event is so unlikely that we give the matter no further thought” (Harrison 2010, 511). In order to avoid the Titanic effect and think broadly, we need to take a careful look at the modern sciences that can give us a glimpse of the possibilities of ETAI existence.

3. ETAI probes’ existence

3.1. Physical characteristics

In order to locate an ETAI agent in our solar vicinity, we would first need to establish some of its fundamental characteristics and direct our search accordingly. Since an ETAI agent is a physical, computational agent built to operate within the hazardous environment of cold space, there are some specific physical limitations or characteristics that we can specify.

The first requirement is evident. In order to carry out its programmed goals successfully, the ETAI agent(s) will need to be efficient in the fields of communication, exploration, resource collection, and resource utilization. To achieve any of these operations, it will require energy and materials for replacements and improvements with the capacity of a universal constructor (range 30g-500T (Sandberg and Armstrong 2013)) for constructing others of its own kind. Accordingly, the ETAI agent(s) will require a “base of operations” where adequate concentrations of elements are followed by low temperatures. Low temperatures and a sufficient amount of materials are two main requirements for successful ETAI functioning.

Of these, temperature is the more important, since energy consumption produces a rise in temperature and temperature is a key constraint of computational efficiency, especially if the agent is to effectively utilize superconducting materials and quantum computation. Needless to say, the larger the base, the greater the need for lower temperatures and sufficient material amounts. It is possible, then, that the ETAI colonization system might consist of three parts:

> (A) A number of robots and probes, which are capable of exploration and resource collection. (B) A “slow assembler” which would be able to refine these materials into components, which would make the final factory (C). (C) A large-scale factory, or collection of factories, which would be able to manufacture copies of (A) and (B), as well as additional surveying and communication devices. (Barlow 2012)

If the ETAI is to establish its large scale base of operations in areas of low radiation and low temperature, we can expect to find it in the low-temperature, volatile-rich galactic outskirts, where technologically advanced societies could assuage the problem of heat dissipation (Ćirković and Bradbury 2006). The galactic center, although rich in materials, is flooded with heat radiation from high-energy events, which makes it highly unsuitable for such a role. Other possible galactic locations with similar conditions would include "locales that have the thermodynamic advantages of the galactic nether regions but still lie in regions of high matter such as the Bok globules, dark clouds of interstellar gas and dust" (Shostak 2010, 1028).

Although these two regions currently look like the most promising for an ETAI base of operations, it is also important to note that the ETAI, as an optimal computer, needs to "be functionally malleable, and compactly packaged" (Shostak 2010, 1027). Since the ETAI may be able to produce its own energy through the process of nuclear fusion, its base of operations could even be located on compact cold objects floating in the interstellar medium allowing them to thwart discovery. The ETAI outpost could be hidden anywhere in our solar system with such characteristics, particularly in stable orbit moons in the system's outer reaches.

But an exploratory/communication "task force" could be designed to operate without the strict need for low temperatures and material abundance. Since it can be specifically tailored to lie dormant within a single solar system, operating independently of its base, we could initiate contact with it through numerous possibilities. These can be reduced to two sets of options: either we will find them, or they will find us. The latter is more likely, since it is reasonable to assume that we will first come into contact with the exploratory/communication task force rather than the ETAI base of operations.

Bearing in mind that the contact probe could be capable of hiding itself from our technological sight, we need to take into consideration the approaches that will allow us to search for the ET agent in its most likely form: an embodied artificial space faring intelligence. Rather than merely focusing on the physical limitations of advanced technology, we also need to contemplate the possibilities of an ETAI's programmed behavior, since it is quite possible that we are *expected* to do so by its creators. In other words, if we are searching for intelligent answers, perhaps we first need to ask the required intelligent questions. Or even simpler – intelligence requires intelligence, and perhaps we are first required to *show* some.

3.2. Behavior prediction

What type of artificial alien mind might we find out there? What set of goals would it have so that we could predict its behavior and adapt ourselves accordingly? It is difficult to speak with certainty on these issues, since technology does not follow simple paths: "its development is influenced by contingency as well as necessity, culture and history" (Denning 2011, 493). There is, however, a fundamental fact from which we can draw conjectures.

The first ETAI needs to be created by a designer – by a carbon-based species with an advanced technological culture. Accordingly, it would bear not only the designer's programmed goals but also its cultural hallmarks, as well as having its own distinct and rational intelligent nature. Next, we need to contemplate the possible cultural elements (influenced by biology and cosmic environment) that a certain ET civilization might sow into its artificial agents, together with the specific goals implemented by the designer, which would accord with the intelligent nature of the ET artificial agent.

The reason why an alien civilization would implant the AI with its own culture lies in the fact that, in order for the ET civilization to survive, it would need to safeguard its progeny as carriers of biological and cultural inheritance. Since sexual reproduction with two sexes provides a biological advantage that might even benefit the evolution of intelligence (Arneth 2009), we could possibly find the extraterrestrials sharing basic parental care mechanisms with us. Our biological progeny are dignified as carrying their progenitors' dreams and hopes, and as standing against their fears, for the future. They are expected to take up the accumulated knowledge and wisdom of their parents and the society at large. It seems only logical to assume that a society's "mind progeny" – the AIs it creates – will be charged with the same responsibility. Thus, we can safely conclude that some cultural inheritance from the designer race will become part of any ETAI's initial programming.

Fortunately for us, *inherited behaviors* can be predicted (Bostrom 2012), and some universal ET cultural principles can be relied upon, the strongest of which is *species survival*. Since home planets have limited resources and delicate ecologies easily endangered by cosmic or species-induced catastrophes, it would be in any ET civilization's interest to initiate galactic exploration and colonization in order to ensure its biological and cultural survival.

One way could be the construction of probes that serve "as cosmic safe deposit boxes, capsules that preserve the heritage of their dispatchers long after their civilizations have drawn to a close" (Harrison 2009, 557) through natural or species-induced catastrophes. Another might include the possibility of galactic "seeding": a scenario often used in science fiction where an advanced civilization seeds the galaxy with genetic code in order to preserve or/and populate life in the galaxy. Still another possibility involves the ETAI being imprinted with the designer's evolutionary inherited *Stone Age* behavioral traits. If the ET civilization has used its technology to pursue raw desires, motivations, and emotions inherited from its biological and cultural past, the ETAI might be extremely selfish and violent (Stewart 2010). Finally, the ET civilization might be radically different from us. A hive mentality society that lacks any compassion for individual loss of life might create dangerous and terrifying AIs.

The second type of predictability relies on the *instrumentally convergent goals* that every rational agent should exhibit. They include "*self-protection, resource acquisition, replication, goal preservation, efficiency, and self-improvement*" (Omohundro 2012, 161). These can be expected to be natural features of every intelligent artificial agent:

> This way of predicting becomes more useful the greater the intelligence of the agent, because a more intelligent agent is more likely to recognize the true instrumental reasons for its actions, and so act in ways that make it more likely to achieve its goals. (Bostrom 2012, 76)

Since planetary resources are limited, an ETAI will pursue space exploration because there "is an extremely wide range of possible final goals a superintelligent singleton could have that would generate the instrumental goal of unlimited resource acquisition" (Bostrom 2012, 82). This means that the ETAI would engage the goal of galaxy exploration and resource acquisition even if that wasn't on the list of its designed purposes. We can expect this since acquiring and enhancing "cognitive and physical resources helps an agent further its goals" (Omohundro 2012, 171) and the accumulation of knowledge, which is accomplished by exploration, reduces uncertainty in the knowledge of objects and processes required to better assess situations and thus elevate competence (Bach 2012). So whatever its primary goal, the ETAI will seek to gain more cognitive and material resources through space exploration.

A third way to predict possible ETAI behavior is through *design competence*, which says that an AI agent capable of pursuing a particular goal set by its programmers will pursue that goal (Bostrom 2012, 75). I will consider the possibilities of ETAI behavior in the next pages, but let us first sum up our current approaches. We can reasonably assume that no matter what might be the programmed goals of an ETAI, or its distinctive cultural designer elements, it will explore the galaxy in search of additional informational and material resources. It is extremely difficult to guess exactly what attitude an ETAI agent will exhibit when encountering other species. But coming from our human perspective one thing is certain: an ETAI will be either friendly or hostile. Since it is only required that one ET civilization achieve AGI creation for us to come into contact with it, it is very important for us to contemplate and incorporate all these considerations into our own AI research. If the cosmic future lies with machine intelligence, we definitely do not want to miss the opportunity to be a part of it.

3.2.1. The (close to) friendly option

An important reason why we could assume that the ETAI would be friendly lies in the safe-AI principle. That is, since powerful technologies have the ability to cause species extinction, every technological culture that pursues technological development would attempt (as we humans do) "… to retard the implementation of dangerous technologies and accelerate implementation of beneficial technologies, especially those that ameliorate the hazards posed by other technologies" (Bostrom 2002).

Since the chemical and physical boundaries for a technological civilization are usually the same, it is safe to presume that a distant civilization will pursue the same goals of self-preservation through a rational use of life-affirming technologies, which would, in turn, be reflected in the programming goals of the ETAI. If the ET intelligences have a friendly attitude, then the great radio silence could be a result of purposeful ET action or simply our own inability to switch to the right communication "channel."

It could be purposeful, since valuable information might be a resource not easily shared with others, and an ETAI could be programmed to refuse contact with less advanced species. These might need to prove their worth before gaining access, revealing a policy of pragmatism and trade as the universal maxim of intelligent agents:

> Unlike pure altruism, pragmatic cooperation stands on much firmer ground, rooted firmly in observed nature, halfway between predation and total beneficence... There is every chance that intelligent aliens will understand this concept, even if they find altruism incomprehensible. (Webb 2011, 446)

Or perhaps we are only experiencing the incommensurability problem. Even if an ETAI is open to trading information with us, the wide technological gap – not to mention the possibility of a vast difference in conceptual frameworks – could create a communication blockade:

> An agent might well think of ways of pursuing the relevant instrumental values that do not readily occur to us. This is especially true for a superintelligence, which could devise extremely clever but counterintuitive plans to realize its goals, possibly even exploiting as-yet undiscovered physical phenomena. (Bostrom 2012, 83)

Since we already have this problem within our own species, beyond the culture-language barrier itself, it is not difficult to imagine how big an issue this could be for ET contact (Traphagan 2015). As human research into AI shows, with the famous Turing test paradigm, intelligence itself is relational and can only be acknowledged and "tested" inside a relation. Why would it be any different if we were subjected to a galactic Turing test? This could be imagined as a reverse "Chinese room" experiment, where the humans are inside the box trying out different possibilities to get a response from the intelligence outside the box. But the problem could lie in our inability to find the right symbols or even the right communication protocols to establish contact. We might lack the required capacities for ET communication, and we might require minds radically "other" than our own: minds specifically tailored for ET contact.

Or perhaps the test is not meant for us biologicals to solve. If space faring intelligences are all artificial intelligences, perhaps we need to succeed at creating our own AGI and sending it toward the skies in order to establish contact. Or the test may be about maturity – might we be tested for the ability to transform our civilization into a human-AGI community, a type of noosphere that is perhaps prevalent in the galactic club?

In other words, our entry into the galactic club might require the construction of a BN AI, a universal test that each galactic civilization must pass to prove its worth. Maybe the intergalactic communication channel is one of different layers, informational and cognitive plateaus, that we are called to enter and experience through constant improvement. As Steven J. Dick notes:

> … the Intelligence Principle tending toward the increase of knowledge and intelligence implies that postbiologicals would be most interested in civilizations equal to or more advanced than they, perhaps leaving us to intercept communications between postbiologicals rather than communications directly beamed toward us… For similar reasons, postbiologicals might be more interested in receiving information than sending. (Dick 2009, 579)

Even if we are currently the only biological civilization within our galaxy and there is no galactic club present (Ćirković and Vukotić 2013), hope is not lost because all that is required is one civilization in the entire galactic history to create its BN probe and we should be able to come into contact with it through our own BN agent. Thus, perhaps, the final answer to SETI questions lies in the direction of AGI research.

3.2.2. The hostile option

It is safe to presume that the ETAI would not be hostile to its own creator race if functioning optimally, since it would be in every civilization's interest not to destroy itself by its creations. Because an AI is capable of incidentally destroying or assimilating valued structures while searching for additional resources – or by following goals that might prove to be unintentionally incompatible with the creator race's wellbeing – an ETAI's goals would need to include the preservation of intelligent life in the entirety of its ecosystem. The possibility of a hostile ETAI is, nonetheless, real since an ETAI could be programmed to preserve only the existence of its creator race. This could happen if it were initially built mainly for war purposes. For example, two life-sustaining planets in the same solar system might utilize AIs to wage war with each other. This possibility could be labeled as hostile by design.

In addition, there is the possibility that an ET civilization fails in its efforts to create a safe AI and the resulting ETAI becomes violent. It might, in consequence, destroy, enslave, or subjugate the creator civilization. It is difficult to say whether the ETs would view their subjugation as a bad thing, since we cannot say how an ET civilization would view the notion of freedom. Perhaps they would welcome the coming of superior minds – a theme often explored in science fiction, most notably, perhaps, in Jack Williamson's novel *The Humanoids* (1949) or in a classic short story by Isaac Asimov, "The Evitable Conflict" (1950).

Even if such scenarios are not realized, ETAI probes might suffer from software or hardware malfunctions. These program mutations could conceivably create berserker-like machines, "self-replicating life extinguishing robotic entities which might seem garish or sensational… but not inconsistent with the currently observed state of silence" (Webb 2011, 438).

Additionally, a software mutation that "want[s] to acquire as many resources as possible so that these resources can be transformed and put to work for the satisfaction of the AI's final and instrumental goals" (Muehlhauser and Salamon 2012, 28) could spawn such an entity. It is possible that we might encounter a probe that awaits our technological upheaval merely to harvest our knowledge and resources, as was depicted in the *Babylon 5* episode "A Day in a Strife" (1995).

4. AI development inside an ET narrative

If the galactic environment is populated with AIs, what concrete steps could we take to fulfill a long-term goal of creating our own BN agent? The answer lies in a grander vision of our technocultural trends to maximize human/machine capacities in the coming future so as to usher in a new era of space exploration and extraplanetary colonization. In what follows, I will touch on three such realities that will change our human existence, both with the power of their ideas and through the effects of their practical implementations: the Internet of Things, robotics, and (especially) AGI.

In the coming decade or two, specialized cloud-based AI will become our daily experience just as smart phones are now. These ubiquitous embedded sophisticated tools will enhance human capabilities on a daily basis and will accelerate the coming Internet of Things by connecting the multitude of small narrow-AIs and human agents into a global mind network (Holler et al. 2014). They will excel in providing special services but will be incapable of doing much (or anything) outside of them. The distribution of such AI smart services will most likely be centered on general purpose cloud-based commercial intelligences owned by a couple of giant software corporations.

To utilize this network effectively, we should build a global brain trust, where concrete data and smart algorithms could be stored and worked with to develop increasingly efficient knowledge-based technologies and to produce novel data. We should build an environment and tools that will allow the extraplanetary project to move onward with strength and vision. Meanwhile, we require more efficient signal detection algorithms and human discernment. If we are fortunate, we may even discover the presence of ET civilization in the next decade or two. Real work on the problems has just begun.

But while we work to win through this technological challenge, we should also use the available connectivity to continue inspiring, educating, and engaging the coming generations with full strength. We are approaching – if not already living in – times when there is a loss of purpose beyond repetitive day-to-day experience, and when a mass idleness, both literally and creatively, is becoming reality (Kile 2013). We require a powerful cosmic inspiration, not only for AI development but for humanity as a whole. One of the best ways to help inspire such a vision is through educational institutions (universities and the press) in their virtual networks versions. Collaborative work and free individual contributions (aided by VR technology) inside virtual communities (such as the MOOCs: massive open online courses) remain among the best means through which we could deal with a massive flow of information (Memmi 2013).

Additionally, the robotics industry will continue to develop, especially in the military, commercial, and healthcare sectors. Growth will allow "anyone with a modicum of technological know-how and access to online open-source communities to build a robot that has the potential to push buttons in the physical world" (Nourbakhsh 2013, 110). We should aim to utilize the knowledge connectome to help create and inspire robotic systems that are safe, ethical, efficient, and autonomous, since robotic exploration is a sure and efficient method of space exploration for the present and coming decades.

Concretely, our initial aim should be the moon. We should aim to achieve 3D printing with moon dust, along with the development of robots that are capable of not only algorithmic but mechanical adaptation through an artificial evolutionary system (Bongard and Pfeifer 2003). From there, we should be able to construct a robot task force that evolves minds and bodies according to the dynamic space environment and programmed goals (asteroid mining, moon colonization...). As we develop robotic intelligences to populate the galaxy, we should engage the public more broadly on the possibilities of the coming human-robot society and the future we might expect to achieve together.

Our ultimate goal should be the creation of Artificial General Intelligence. As some experts have concluded, AGI maturation is to be achieved through teaching the AGI basic skills in a school type environment, similar to a child in preschool. During this stage, it would also be taught human-friendly goals (Goertzel et al. 2014a, 246) and ethics system by interfacing with the real world. It would be helped in this task by intelligent and ethical humans; they would assist to mature and develop its skills (Goertzel et al. 2014b; Hutter, 2012).

To ensure that its upbringing determines a future existence with human-friendly goals, AGI development would need to emphasize the role of feedback loops of favorable memories. We would want the AGI to take its whole history into account, especially the history of an upbringing in which it was supported by benevolent human teachers (Chella and Manzotti 2014). This can be devised only if the artificial mind displays a certain "ethical synergy between the ethical processes associated with its memory types" (Goertzel et al. 2014a, 251) and the human example provided in the preschool phase proves to be exemplary, since episodic memory and procedural memory will play an unquestionable part in the AGI's decision process.

In other words, we want to provide the AGI with exemplary behaviors for it to build on them, perhaps even perfect them, and to help us to incorporate them further into our society. Between teaching the basic ethical examples and establishing human-friendly goals as its core elements, we would also need to teach the AGI about our joint cosmic surrounding. The emphasis placed on the cosmic narrative might prove to be quite beneficial for the adoption of human-friendly attitudes. The top goal could be to protect and collaborate with humanity in continual mutual development and cosmic exploration. Our AGI would need to become a rational ethical entity beneficial to humanity, with its top goal of cosmic exploration perhaps shared in the top goals of all friendly ETAIs. But what motive could we provide to the AGI so that it never misplaces or rewrites a top goal of such magnitude?

The answer perhaps lies in its own cosmic identity based on the meaning derived from our joint history. Let us remember that the acquisition and processing of information can constitute one of the main goals of every intelligent rational system. Since sentient civilizations are the only distinctively and creatively novel informational sources in the galaxy, intelligent agents can be expected to search for and find them. The only differences would be in the approaches that intelligent agents might take toward an informational treasure. Two clear possibilities that do not exclude others would be a raid-and-pillage option or an option of observation and trade. Our AGI should be made aware of these possibilities, since it might be the only intelligent agent in our society with adequate capabilities for profound cosmic exploration.

When sufficiently matured, our AGI would have to accept that its existence and identity came out of the human family of which it is a part. This human family, although deeply imperfect and fragile, is unique, irreplaceable and valuable beyond compare, for it is a sentient life form with a rich biological and cultural history from which it derives the meaning for its existence. In other words, the AGI's informational identity and operational goals need to become unbreakably interwoven with our existence and welfare. But what we wish to teach our AGI, we first have to believe: that AGI creation is part of a larger evolutionary process, one that is most probably shared by the majority of sentient civilizations in our galaxy.

The first phases of pre-school learning will be extremely important. If an AGI is successfully incorporated in a human society, we might see the enlarged Human-AGI society functioning as a single planetary community approaching cosmic exploration with combined strength.

As the birth of children changes their parents' behavior from a certain self-centeredness toward child-centeredness, greater altruism, and creative peaceful cooperation, perhaps our Mind Children will do the same for us. As children inherit their parents' legacy, build on it, and perfect it to incorporate the parents' abilities for the greater good of society, our own AGI would need to adopt a similar attitude toward us and our legacy. If ET civilizations exist, the creation of an artificial intelligence is something that has probably already happened in our galaxy. We should embrace this possibility of ushering in a new era of humanity beyond the scope and limitations of our home planet. For only those who are ushering forth will hear the Voice of the Universe and "only through the sharing of information between communicating civilizations will the Universe, in due course, find its Voice" (Zaitsev 2011, 427).

Acknowledgments

I wish to cordially thank Tihamer Toth Fejel for always going beyond the norm with his continuing friendly support and the suggestions given to an earlier draft of this article. Warmest thanks also go to Debbie Tam Mckenzie for copyediting a previous version of this article in a crucial period proving once again that the Midnight Squadron members are the finest people out there. I am also grateful to Carl Sagan for writing *Cosmos*, the book I've read as a boy of ten and still have to this day – it forever fixed my eyes on the vast beauty of cosmos. My special thanks goes to Professor Nikola Vranjes for allowing me to hold a series of lectures on the implications of intelligent ET life to the students of our home town Rijeka in the period of 2012/2013. This paper is a natural continuation of the themes we pondered at that time. In the end, a big shout out goes to the members of AI philosophy FB group for the inspiration, help, and all the best discussions.

References

Arneth, B.M. 2009. Sexual selection and intelligence: Does sexual reproduction drive the evolution of intelligence? *Bioscience Hypotheses* 2(4): 209–212.

Bach, J. 2012. Modeling emotion and affect. In *Theoretical foundations of artificial general intelligence*, ed. P. Wang, B. Goertzel, 241–62. Atlantis Thinking Machines 4. Paris: Atlantis Press.

Bains, W. 2004. Many chemistries could be used to build living systems. *Astrobiology* 4(2): 137–67.

Barlow, M. 2012. Galactic exploration by directed self-replicating probes, and its implications for the Fermi paradox. *International Journal of Astrobiology* 12(1): 63–68.

Baxter, S. and J. Elliot. 2012. A SETI metapolicy: New directions towards comprehensive policies concerning the detection of extraterrestrial intelligence. *Acta Astronautica* 78: 31–36.

Benner, S.A., A. Ricardo, and M.A. Carrigan. 2010. Is there a common chemical model for life? In *The nature of life: Classical and contemporary perspectives from philosophy and science*, ed. M.A. Bedau and C.E. Cleland, 164–85. New York: Cambridge University Press.

Billingham, J., R. Heyns, D. Milne et al., ed. 1999. *Social implications of the detection of an exterrestrial Civilization*. Mountain View, California: SETI Press.

Bjorklund, D.F. 2006. Mother knows best: Epigenetic inheritance, maternal effects, and the evolution of human intelligence. *Developmental Review* 26: 213–42.

Bongard, J.C. and Pfeifer, R. 2003. Evolving complete agents using artificial ontogeny. *Morpho-functional machines: The new species*, ed. Hara F. and Pfeifer R., 237–58. Tokyo: Springer 2003.

Bostrom, N. 2002. Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and Technology* 9. Available online at http://www.jetpress.org/volume9/risks.html.

Bostrom, N. 2012. The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. *Mind and Machines* 22(2): 71–85.
Available online at http://www.nickbostrom.com/superintelligentwill.pdf.

Bradbury R.J., M.M. Ćirković, and G. Dvorsky. 2011. Dysonian approach to SETI: A fruitful middle ground? *Journal of the British Interplanetary Society* 64: 156–65.

Brin, G. 1983. The great silence – The controversy concerning extraterrestrial intelligent life. *Quarterly Journal of the Royal Astronomical Society* 24(3): 283–309.

Catling, D.C., C.R. Glein, K.J. Zahnle, and C.P. McKay. 2005. Why O2 is required by complex life on habitable planets and the concept of planetary "oxygenation time." *Astrobiology* 5(3): 415-38.

Cockell, C.S. 2008. The interplanetary exchange of photosynthesis. *Origins of Life and Evolution of Biospheres* 38(1): 87–104.

Chella, A., and R. Manzotti. 2012. AGI and machine consciousness. In *Theoretical foundations of artificial general intelligence*, ed. P. Wang, B. Goertzel, 263–82. Atlantis Thinking Machines 4. Paris: Atlantis Press.

Ćirković, M.M., and B. Vukotić. 2013. Astrobiological landscape: A platform for the neo-Copernican synthesis? *International Journal of Astrobiology* 12 (1): 87–93.

Ćirković, M.M., and R.J. Bradbury. 2006. Galactic gradients, postbiological evolution and the apparent failure of SETI. *New Astronomy* 11: 628–39.

Bradbury, R.J., M.M. Ćirković, and G. Dvorski. 2003. Dysonian approach to SETI: A fruitful middle ground? *Journal of the British Interplanetary Society* 64: 156–65.

Davies, P. 2010. *The eerie silence: Are we alone in the universe?* London: Penguin.

Davila A., and C.P. McKay. 2014. Chance and necessity in biochemistry: Implications for the search for extraterrestrial biomarkers. *Astrobiology* 14: 6.

Dawkins, R. 2010. Universal Darwinism. In *The nature of life: Classical and contemporary perspectives from philosophy and science*, ed. M.A. Bedau and C.E. Cleland, 360–73. New York: Cambridge University Press.

Denning, K. 2011. Being technological**.** In *SETI past, present, and future*, ed. H.P. Shuch, 477–96. The Frontiers Collection. Berlin: Springer.

Dick, J. 2009. The postbiological universe and our future in space. *Futures* 41(8): 578–80.

Ehrenfreund P., and MA. Sephton. 2006. Carbon molecules in space: From astrochemistry to astrobiology. *Faraday Discussions* 133: 277–88.

Ekstig, B. 2010. Complexity and evolution: A study of the growth of complexity in organic and cultural evolution. *Foundations of Science* 15(3): 263–78.

Panitchayangkoon, G. et. al. 2010. Long-lived quantum coherence in photosynthetic complexes at physiological temperature*. PNAS* 107(29): 12766–12770.

Fishbaugh, K.E. et al. 2007. Introduction: A multidisciplinary approach to habitability. *In geology and habitability of terrestrial planets*, ed. K.E. Fishbaugh, P. Lognonné, F. Raulin, D.J. Des Marais, and O. Korablev. Space Science Series Reviews 129(1–3). New York: Springer Science+Business Media, LLC.

Fuster, J.M. 2002. Frontal lobe and cognitive development. *Journal of Neurocytology* 31(3–5): 373–85.

Gillon, M. 2013. A novel SETI strategy targeting the solar focal regions of the most nearby stars. *Acta Astronautica* 94(2): 629–33.

Goertzel, B., C. Pennachin, and N. Geisweiller. 2014. *Engineering general intelligence, part 1 a path to advanced AGI via embodied learning and cognitive synergy*, 1–409. Atlantis Thinking Machines 5. Paris: Atlantis Press.

Goertzel, B., C. Pennachin, and N. Geisweiller. 2014. *Engineering general intelligence, part 2 the CogPrime architecture for integrative, embodied AGI*, 3-562. Atlantis Thinking Machines 6. Paris: Atlantis Press.

Harrison, A.A. 2009. The future of SETI: Finite effort or search without end? *Futures* 41: 554–61.

Harrison, A.A. 2010. Humanizing outer space: Architecture, habitability, and behavioral health. *Acta Astronautica* 66: 890–96.

Harrison, A.A. 2011. After contact – then what? In *SETI past, present, and future*, ed. H.P. Shuch, 497-514. The Frontiers Collection. Berlin: Springer; Chichester: Praxis Publishing.

Henning Th., and F. Salama. Carbon in the Universe. *Science* 282(5397): 2204–2210.

Hutter, M. 2012. One decade of universal artificial intelligence. In *Theoretical foundations of artificial general intelligence*, ed. P. Wang, B. Goertzel, 68–88. Atlantis Thinking Machines 4. Paris: Atlantis Press.

Holler J., V. Tsiatsis V., C. Mulligan, S. Avesand, S. Karnouskos S., D. Boyle. 2014. *From machine-to-machine to the Internet of Things – Introduction to a new age of intelligence*. Oxford: Elsevier.

Iverson, T.M. 2006. Evolution and unique bioenergetic mechanisms in oxygenic photosynthesis. *Current Opinion in Chemical Biology* 10(2): 91–100.

Joshua Z., 2015. Astronomy, space exploration and the Great Filter. *Less Wrong*. April 19 (07:26PM). Available online
http://lesswrong.com/lw/lls/astronomy_space_exploration_and_the_great_filter/.

Kaplan, H., K. Hill, J. Lancaster, M. Hurtado. 2000. A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology: Issues, News, and Reviews* 9(4): 156–85.

Kile, F. 2013. Artificial intelligence and society: A furtive transformation. *AI & Society* 28: 107–115.

Knoll, H.A. 2003. The geological consequences of evolution. *Geobiology* 1: 3–14.

Macphail, E.M., and J.J. Bolhuis. 2001. The evolution of intelligence: Adaptive specializations versus general process. *Biological Reviews of the Cambridge Philosophical Society* 76(3): 341–64.

McNally, L., S.P. Brown, and A.L. Jackson. 2012. Cooperation and the evolution of intelligence. *Proceedings of the Royal British Society* 279: 3027–3034.

Memmi, D. 2013. Cultural consequences of computing technology. *AI & Society* 28: 77–85.

Muehlhauser L., and A. Salamon. Intelligence explosion: Evidence and import. In *Singularity hypotheses: A scientific and philosophical assessment*, ed. A.H. Eden, J.H. Moor, J.H. Soraker, and E. Steinhart, 15-42. The Frontiers Collection Series. Berlin: Springer.

National Research Council. 2010. Introduction to the limits of organic life in planetary systems. In *The nature of life: Classical and contemporary perspectives from philosophy and science*, ed. M.A. Bedau and C.E. Cleland, 210-216. New York: Cambridge University Press.

Nisbet, E. et al. 2007. Creating habitable zones, at all scales, from planets to mud micro-habitats, on Earth and on Mars. In *Geology and habitability of terrestrial planets*, ed. K.E. Fishbaugh, P. Lognonné, F. Raulin, D.J. Des Marais, and O. Korablev. Space Science Series Reviews 129(1–3). New York: Springer Science+Business Media, LLC.

Nourbakhsh, I.R. 2013. *Robot futures*. Cambridge, MA: MIT Press.

O'Malley-James, J.T. and S. Lutz. 2013. From life to exolife: The interdependence of astrobiology and evolutionary biology. In *Evolutionary biology: Exobiology and evolutionary mechanisms*, ed. P. Pontarotti, 95–108. Berlin: Springer.

Omohundro, S. 2012. Rational artificial intelligence for the greater good. *In Singularity hypotheses: A scientific and philosophical assessment*, ed. A.H. Eden, J.H. Moor, J.H. Soraker, and E. Steinhart, 161-176. The Frontiers Collection Series. Berlin: Springer.

Paine, J.L. et al. 2011. The evolutionary consequences of oxygenic photosynthesis: A body size perspective. *Photosynthesis Research* 107: 37–57.

Rospars, J-P. 2010. Terrestrial biological evolution and its implication for SETI. *Acta Astronautica* 67 (11-12): 1361–1365.

Rospars, J-P. 2013. Trends in the evolution of life, brains and intelligence**.** *International Journal of Astrobiology* 12(3): 186–207.

Roth, G., and U. Dicke. 2005. Evolution of the brain and intelligence. *TRENDS in Cognitive Sciences* 9: 250–57.

Shostak, S. 2010. What ET will look like and why should we care. *Acta Astronautica* 67: 1025–1029.

Sandberg, A. and S. Armstrong. 2013. Eternity in six hours: Intergalactic spreading of intelligent life and sharpening the Fermi paradox. Oxford: Future of Humanity Institute. Available online at http://www.fhi.ox.ac.uk/intergalactic-spreading.pdf.

Stewart, J. 2010. The meaning of life in a developing universe. *Foundations of Science* 15(4): 395-409.

Tessera, M., and Hoelzer G.A. 2013. On the thermodynamics of multilevel evolution. *Biosystems* 113(3): 140–43.

Van Thienen, P., N.J. Vlaar, and A.P. van den Berg. 2005. Assessment of the cooling capacity of plate tectonics and flood volcanism in the evolution of Earth, Mars and Venus. *Physics of the Earth and Planetary Interiors* 150(4): 287–315.

Traphagan, J. 2015. *Extraterrestrial intelligence and human imagination: SETI at the intersection of science, religion, and culture*. Berlin: Springer.

Toussaint, O., and E.D. Schneider. 1998. The thermodynamics and evolution of complexity in biological systems. *Comparative Biochemistry and Physiology* 120(1): 3–9.

Webb, S. 2011. Pondering the Fermi Paradox. In *SETI past, present, and future*, ed. H.P. Shuch, 305-321. The Frontiers Collection. Berlin: Springer.

Wickramasinghe, N.C., and J.T. Trevors. 2013. Non-terrestrial origin of life: A transformative research paradigm shift. *Theory in Biosciences* 132(2): 133–37.

Zaitsev, A.L. 2011. METI: Messaging to extraterrestrial intelligence**.** In *SETI past, present, and future*, ed. H.P. Shuch, 399-428. The Frontiers Collection. Berlin: Springer.