

That Which Withholds Eschaton: The Katechon Hypothesis for Resolving the Fermi Paradox

Abstract: A corollary of the Simulation Argument is that the universe's computational capacity may be limited. Consequently, advanced alien civilizations may have incentives to avoid space colonization to avoid taking up too much "calculating space" and forcing a simulation shutdown. A possible solution to the Fermi Paradox is that analogous considerations may drive them to avoid broadcasting their presence to the cosmos, and to attempt to destroy or permanently cripple emerging civilizations on sight. This game-theoretical equilibrium could be interpreted as the "katechon" - that which withholds eschaton - doom, oblivion, the end of the world. The resulting state of mutually assured xenocide would result in a dark, seemingly empty universe intermittently populated by small, isolationist "hermit" civilizations.

Keywords: aliens; digital physics; ETIs; existential risks; Fermi Paradox; metaphysics; simulation hypothesis.

Authors: Anatoly Karlin

Date: October 21, 2019

Table of Contents

Table of Contents	2
Introduction	3
Filtering the Great Filter	4
Shadows of the Past	5
Gloomy Presentiments	7
Universe Simulations	15
Calculating Space	16
Computing Costs of Cosmic Expansion	18
The Katechon Hypothesis	24
Do Aliens Believe in The Matrix?	25
Can Expansion Be Controlled?	28
Space Sakoku and Zero-Sum Cosmopolitics	33
The RMAX Equilibrium	35
Navigating the Black Seas of Infinity	39
Acknowledgements	43
Bibliography	44

Introduction

In October 2017, a strange object appeared in the skies. ‘Oumuamua, or “scout” in Hawaiian, was the first confirmed interstellar object to pass through our Solar System [1]. As Robin Hanson pointed out, it was “suspiciously extreme in many ways”: Highly elongated, with a very fast rotation speed, no outgassing as with comets, and a strongly red color typical of metallic asteroids¹. Could it actually have been a “scout” in the most literal sense of the word? The suggestions that it might be an alien spaceship did not just come from hyperactive sci-fi aficionados².

The recent discovery of the more typical 2I/Borisov suggests that interstellar visitors are far more common than previously thought. Nonetheless, this doesn’t contradict the possibility that one fine day in the coming century, one such “scout” from outer space will do in our civilization and our species, “with no warning and for no apparent reason” (with due apologies to Neal Stephenson). As I will argue in this article, this reason may well be not only perfectly logical, but born out of existential necessity.

¹ Robin Hanson’s Twitter: <https://twitter.com/robinhanson/status/936769317349347329>

² For instance, “Is 1I/‘Oumuamua an Alien Spacecraft?” by Penn State astrophysicist Jason T. Wright: <http://sites.psu.edu/astrowright/2017/11/22/is-1-ioumuamua-an-alien-spacecraft/>

Filtering the Great Filter

The most merciful thing in the world, I think, is the inability of the human mind to correlate all its contents. We live on a placid island of ignorance in the midst of black seas of infinity, and it was not meant that we should voyage far. The sciences, each straining in its own direction, have hitherto harmed us little; but some day the piecing together of dissociated knowledge will open up such terrifying vistas of reality, and of our frightful position therein, that we shall either go mad from the revelation or flee from the deadly light into the peace and safety of a new dark age. – H. P. Lovecraft.

Robin Hanson's answer to the Fermi Paradox - "where is everyone?" - is that the apparent rarity of advanced alien civilizations is due to some bottleneck event that all intelligent life has to go through [2]. One may view this concept as an extension of the Drake Equation, under the additional assumption that some of its values are so low that the average galaxy isn't likely to host much more than one civilization that emits detectable signals into space at any particular point in time.

It is possible that the Great Filter lies in our past, meaning that we are safe, and a ball of hedonium may soon envelope our planet and suffuse our future lightcone. In a recent paper, a team of futurists recalculated the Drake Equation; instead of using point estimates, which typically yield an infinitesimal probability of our galaxy containing no alien civilizations, they calculated the distribution of expert probability estimates, ran a Monte Carlo simulation, and deduced that there is a one in three chance that we are alone in our galaxy, effectively "dissolving" the Fermi Paradox [3]. We should hope that they are correct, since the alternative possibility - that the Great Filter lies in our future - very likely dooms us to imminent extinction.

Shadows of the Past

If the Great Filter lies in our past, then it would imply that at least one of the former is very rare or improbable:

- The evolution of life
- Certain evolutionary milestones
- The evolution of intelligence
- Advanced technological civilization

The evolution of life. Straddling the warm “Goldilocks zone” between the Sun and the cold emptiness of outer space, perhaps Earth was uniquely optimal for the emergence of life [4]. This “Rare Earth Hypothesis” has been challenged in recent years by the discovery of vast numbers of Earth-like planets. However, perhaps a “weak” version of REH might still hold - that Earth is optimal for the *fast* emergence of complex, intelligent life. For instance, at least two critical evolutionary leaps - the appearance of eukaryotes, and of large metazoans - may have depended on large, creative-destructive oxygenation events [5]. If the Earth’s oxygen-absorption capacity had been higher, complex life might not have had time to emerge by the time the Sun fried our planet in another billion years.

Evolutionary milestones. Life appeared - in geological terms - almost immediately after the formation of Earth [6]. So abiogenesis is unlikely to have been the principal barrier. Following the evolution of the first self-replicating molecules, there was 1.8 billion years of near biological stasis [7]. Perhaps the likeliest candidate for the Great Filter was the transition from prokaryotes to eukaryotes, which were a prerequisite for the appearance of complex, multicellular organisms and sexual reproduction (both of which may have also been improbable). Conversely, transitions that evolved independently on many separate occasions - limbs, sight, photosynthesis - may be safely ruled out.

The evolution of intelligence. Nervous systems with distinct neural signaling molecules - the building blocks of big brains and intelligence - evolved

independently across both ctenophores and cnidarians/bilaterians half a billion years ago [8]. Moreover, according to the Red Queen hypothesis, organisms don't exist in a vacuum, but need to compete against other organisms within a mutable environment [9]. Since you can't stand still indefinitely, this should drive the evolution of more complex lifeforms. This theory is backed by evolutionary history; since the Cambrian Explosion 550 million years ago, the maximum encephalization rate has been constantly doubling every 50 million years [10]. More broadly, there has been exponential growth in minimum genome size since the dawn of life [11]. As Pierre Teilhard de Chardin and Vladimir Vernadsky had intuited in the 1920s, evolution is "creative", with an overarching teleological drive towards greater complexity and intelligence.

Nor is there good reason to believe that there is anything particularly unique or improbable about *human* intelligence. The world after the dinosaurs has seen convergent evolution of greater intelligence across the entire swathe of the world's habitats and animal taxa - dolphins, whales, and octopuses in the oceans; corvids and Gray African parrots in the skies; and the Great Apes, elephants, and some monkey species on land³. If humans were to disappear off the face of the planet today, there will be plenty of candidate species to rekindle civilization. For our own part, hominins only split off from primates 15 million years ago, and our brains have been exploding in size and capability ever since. Even the theory that there was a discrete "great leap forward" in human behavioral modernity 50,000 years ago has been sidelined in favor of explanations stressing continuous and accelerating change [12].

Advanced technological civilization. Technological growth has been increasing at an exponential rate with remarkable consistency for mankind's entire history [13]. Agriculture began 10,000 years ago; the Industrial Revolution took off around 1780. In the past couple of decades, the new science of cliodynamics has provided a strong theoretical basis for this pattern [14]. The basic idea is that as population rises, you get more potential

³ Alex K. Chen has a comprehensive and well-researched, if not entirely rigorous, list of animals ordered by estimated IQ:

<https://www.quora.com/What-is-a-good-list-of-animals-ordered-by-intelligence/answer/Alex-K-Chen>

inventors, who create more technology and increase the carrying capacity of the land, resulting in higher populations, more potential inventors, etc. Although this basic mechanism is punctuated by “Malthusian cycles” - a euphemism for population collapses in the wake of disasters such as droughts, famines, plagues, civil wars, and nomad invasions - the exponential trend was remarkably stable in the long-run [15]. Other positive feedback loops include literacy and “technologies to create more technologies”, such as paper, reading glasses, the printing press, and computers [16]. Human accomplishment, as proxied by the per capita incidence of great scientists and artists, also rises exponentially over the past 2,500 years, peaking in the late 19th century [17].

The history of science and technology is also replete with examples of convergent evolution. Fire was invented, forgotten, and reinvented by countless numbers of disparate human bands [18]. Both agriculture and literacy were independently discovered across multiple locations⁴. The Ancient Greeks almost got to the steam-engine, and there were proto-industrial revolutions in the Roman Empire and Song China before the real deal got going in late 19th century Britain [19]. At that point, the Scientific Revolution had been ongoing for more than two centuries, and more than half the denizens of the European core were literate [20]. At that point, even if Britain had been swallowed up by the sea, an industrial revolution in that region of the world had long since become inevitable.

Gloomy Presentiments

Future Great Filters are *mostly* coterminous with the concept of “existential risks” and determine the value of the final term in the Drake Equation, which

⁴ Even if agriculture was impossible in our world, it may not have closed off the road to industrialism. Fishing was able to support large sedentary populations, and even a nomadic existence in the Arctic was not necessarily incompatible with sustained technological progress and the development of industries, e.g., bone armor was already getting manufactured in Siberia 3,900 years ago (see <https://siberiantimes.com/science/casestudy/features/warriors-3900-year-old-suit-of-bone-armor-uncearthed-in-omsk/>).

refers to the “length of time over which [advanced] civilizations release detectable signals.” In a seminal paper, Nick Bostrom defined x-risks thus [21]: “One where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential.” He also pointed out that a very strong indication that the Great Filter lies in our future would be the discovery of evidence of past life - especially complex life - on Mars, or elsewhere in the Solar System. At a single stroke, this would rule out the earliest and some of the strongest candidates for the Great Filter at a single stroke, and would constitute “by far the worst news ever printed on a newspaper cover” [22].

Major candidates for future Great Filters can be divided into three major bins:

- Geoplanetary, e.g. asteroid impacts, Gamma Ray Bursts (GRBs), geomagnetic reversals, supervolcano eruptions, megatsunamis.
- Technogenic, e.g. nuclear warfare, anthropogenic climate change, pollution, resource depletion, dysgenic decline, bioweapons, nanotech, malevolent AGI.
- Theoretical, e.g. “superpredator” civilizations, “our simulation shuts down.”

Geoplanetary X-Risks. These are characterized by being either eminently survivable - at least on a civilizational scale - or highly unlikely even on terran timescales. One study found that there's a <1% chance that our galaxy will produce a GRB *anytime* in the future, let alone one pointed at the Earth in the near future [23]. If our hunter-gatherer ancestors survived the Toba eruption 72,000 years ago, then modern civilization will surely survive another Yellowstone eruption. An asteroid impact on the level of the K/T extinction event, which did in the dinosaurs, is only expected to happen once every 100 million years [24]. No dangerously large asteroids have been observed on a collision course with Earth, and even if they were... well, if early mammals survived K-T, then so will humans in deep bunkers and nuclear submarines. All of these scenarios will be extremely disruptive, causing dozens if not

thousands of megadeaths. But in almost all these scenarios, some humans will survive, and they will be able to rebuild industrial civilization.

It can't be completely excluded that a *really* large asteroid (>100 km in diameter), a GRB burst, or something more exotic hits us. However, it would be extremely strange – not to mention *extremely suspicious* – if a geoplanetary Götterdämmerung was to do us in just 100-200 years after the invention of radio, during a time when we might be at the cusp of a technological singularity. Then again, we wouldn't have long to speculate about the cosmic injustice of this... which may well be the main point.

Technogenic X-Risks. This consists of “classic” 20th century concerns (nuclear war, dysgenics, and resource depletion), today's “hot” topic of manmade climate change, and “futurist” concerns such as GNR (genetics/nanotech/robotics) technologies and machine superintelligence.

The megatonnage of the world's nuclear stockpiles are an order of magnitude lower today than during the height of the Cold War (not to mention six *orders of magnitude* lower than the energy released during the Chicxulub impact that did in the dinosaurs⁵). Even so, many serious assessments even during the Cold War projected that a solid majority of the American population would survive a full nuclear exchange with the USSR [25,26]. Tens of millions would die, and the bulk of the capital stock in the warring nations would be destroyed, but as Herman Kahn might have said, this as his parody in the movie *Dr. Strangelove* said, this would a regrettable but nevertheless distinguishable outcome compared to a true “existential risk.” In the long term, radioactivity will dissipate, the population will recover, and infrastructure will be rebuilt.

In the past decade, climate activism has aroused the same intensity of passion as concerns over nuclear war in an earlier age. But we need to keep things in perspective. The IPCC does not project global warming much greater than 5.0°C by 2100, even under the most pessimistic emissions scenarios. There is

⁵ Nuclear megatonnage peaked at 20 gigatons in the USA (c.1960) and the USSR (c.1975); both powers are now down to less than 1 gigaton. The Chicxulub impact released ~100,000 gigatons.

also a case to be made that moderate global warming may be a net good in terms of crop yields due to greater precipitation and carbon fertilization [27]. However, even the most extreme projections such as the clathrate gun going off and “zones of uninhabitability” spreading across the mid-latitudes are unlikely to translate into James Lovelock’s apocalyptic visions of “breeding pairs” desperately eking out a hardscrabble survival in the Arctic⁶. The Arctic was a lush rainforest when global temperatures were at such elevated levels, and will be able to support advanced civilization. More importantly, there just isn’t enough sequestered carbon to lead to a runaway greenhouse effect that turns our planet into Venus, at least under current levels of solar radiation [28,29]. A relocation of the locus of human civilization towards the Far North is not an existential risk.

Similar considerations apply to pollution and resource depletion, confusing degradation and difficult adjustments with existential risks. First, it is not clear that things are actually getting worse - environmental standards have been soaring in the developed world (e.g. the Thames is now cleaner than it was in the 16th century), and new technologies are constantly opening up access to previously inaccessible resources (e.g. hydraulic fracking). Second, while future energy shocks may still impinge on living standards, there are no grounds to think they will cause long-term economic decline, let alone technological regression back to Stone Age conditions as some of the most alarmist “doomers” have claimed [30]. There are still centuries’ worth of coal and natural gas reserves left on the planet, nuclear and solar power have only been exploited to a small fraction of their potential, and hydropower - which has some of the highest energy returns on energy invested - isn’t going anywhere. Furthermore, we still have a lot of potential fat to cut! Low car ownership and the extinction of budget airlines do not preclude continued radio emissions or rocketry (e.g., see the USSR).

⁶ Interview with James Lovelock in *The Independent* (2006), see <https://www.independent.co.uk/voices/commentators/james-lovelock-the-earth-is-about-to-catch-a-morbid-fever-that-may-last-as-long-as-100000-years-5336856.html>.

Much of the developed world has experienced dysgenic reproduction patterns - duller people having more surviving children than brighter people - for over a century [31–34]. Although this was long masked by IQ gains from better schooling and nutrition, that process seems to be coming to an end [35,36]. Meanwhile, the problems that need to be solved for scientific and technological progress to continue are getting harder, not easier [16]. Since almost all scientific discoveries accrue to a small cognitive elite in the world's rich, high-IQ nations [37,38], this suggests that technological progress may slow to a crawl during this century as the world's remaining "smart fractions" get depleted [39,40] (assuming that there are no abrupt discontinuities in humanity's capacity for collective problem solving, such as genetic IQ augmentation or machine superintelligence). But will this "idiocracy" be permanent? I doubt it. Since fertility preferences are heritable, and ultra-competitive in a post-Malthusian world, we can expect an eventual reversal of the demographic transition [41,42]. This renewed expansion will last until the world hits the carrying capacity of the late industrial economy, ushering in the return of Malthusianism and reasserting the eugenic fertility patterns of the pre-industrial world [43–46]. Consequently, dysgenic decline does not constitute an existential risk. However, it may have the effect of extending the period of time that future humanity will be subject to increased levels of *other* existential risks.

The final major source of technogenic existential risks concerns new technologies - in particular, genetics, nanotechnology, and robotics ("GNR"). In an ideal world, they promise us a utopian future of radically expanded lifespans and abundant material wealth, if not the secular equivalent of transcendence [13]. But GNR technologies may also be the instruments of our demise. Since bioengineering doesn't require an extensive industrial infrastructure, like a nuclear weapons complex, the means to inflict massive damage may be democratized, vastly increasing the probability of devastating pandemics unleashed through bioerror or bioterror⁷. The engineer Eric Drexler

⁷ E.g. see Bill Joy's classic essay in Wired (2000), "Why the Future Doesn't Need Us" <https://www.wired.com/2000/04/joy-2/>. One researcher believes that this effect may wholly explain the Fermi Paradox [47].

has suggested that nano “engines of creation” may go rogue and blanket the planet in a sea of “gray goo” [48]. However, it is hard to imagine an artificial virus that couples a 100% infection *and* mortality rate, while subsequent research suggests that democidal nanomachine swarms will remain in the realm of science fiction [49]. That said, there are many unforeseen pitfalls - future technology is, almost by definition, unpredictable. So it is not impossible that even something that currently seems highly unlikely (e.g. particle accelerator experiments), if not a complete Black Swan, is what will do us in.

Many experts believe that artificial general intelligence (AGI) will appear by the middle of the 21st century [13,50–52]. An artificial general intelligence (AGI) may be able to quickly bootstrap itself into a superintelligence, defined by Nick Bostrom as “any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest” [53]. Especially if this is a “hard” takeoff, the superintelligence will also likely become a singleton, an entity with global hegemony - and woe to us if it decides to convert us all into paperclips [54,55]! Consequently, the “control problem” in AI is widely considered to be the most acute existential risk in futurist circles.

It also happens to be largely irrelevant to this thesis, as it is one of the few existential risks that do *not* also constitute a Great Filter. Getting turned into paperclips will be bad for us, but irrelevant so far as the Fermi Paradox is concerned. Functionally, all that will happen is that superintelligent machines replace humans as the primary agents of the terran noosphere. Even if Skynet ends up killing us all, why would it stay on this planet indefinitely? More precisely, why would each one of the dozens to millions of past Skynets in our galaxy have uniformly decided to stay on their home planet, instead of beaming their presence out into space or physically expanding their dominion at close to the speed of light? Either machine superintelligence inevitably tends to suicide-or-stasis, which seems intuitively unlikely, though impossible to rigorously disprove since we cannot know the mind of a superintelligence

before inventing one [56]; or superintelligences have universally worked out that they shouldn't broadcast or expand into space from first principles.

Assuming that we cannot “dissolve” the Fermi Paradox to be sufficiently confident in our own solitude, and that geoplanetary and technogenic existential risks don't constitute credible Great Filters, what else could possibly explain the “Great Silence” all around us?

Universe Simulations

Anyone who believes exponential growth can go on forever in a finite world is either a madman or an economist. - Kenneth Boulding.

The concept of “superpredator” civilizations that winnow out any civilization that shows its head above the cosmic parapets – to remove potential competitors; for resources; out of pure psychopathy – is a popular sci-fi trope. Perhaps the reason everyone is silent is because killers stalk the star-strewn skies, hiding in the dark voids between worlds.

I do not consider this to be a very plausible explanation, at least so far as the most commonly cited reasons are concerned. Since any such civilization will have many millions if not billions of years of technological advantage, a posthuman space empire is unlikely to see any civilization at humanity’s 21st century technological level as any sort of significant threat. Nor do I think it likely that they hunger for any resource specific to our humble clump of rock. The universe is full of rocks. Maybe they do it just for the hell of it, like the Reapers of the Mass Effect universe?

Still, I don’t think that’s too likely. First, at least within the human species, empathy has tended to increase with literacy and social complexity, so it’s not too obvious that an evil xenocidal race stalking the heavens would constitute the endpoint of sociocultural evolution [57]. (That said, as I will explain later, evolutionary dynamics may favor the emergence of such civilizations). Second, and more importantly, why would they snuff out fledgling space-traveling civilizations from the shadows - leaving open the chance that some particularly paranoid ones slip by their net - instead of openly exerting dominance and saturating the galaxy with their presence?

Calculating Space

But what if the resource in question is something a bit more... esoteric? In a classic paper from 2003, Nick Bostrom argued that at least one of the following propositions is very likely true: That posthuman civilizations don't tend to run "ancestor-simulations"; that we are living in a simulation; or that we will go extinct before reaching a "posthuman" stage [58]. Let us denote these "basement simulators" as the Architect, the constructor of the Matrix world-simulation in the eponymous film. As Bostrom points out, it seems implausible, if not impossible, that there is a near uniform tendency to avoid running ancestor-simulations in the posthuman era.

There are unlikely to be serious hardware constraints on simulating human history up to the present day. Assuming the human brain can perform $\sim 10^{16}$ operations per seconds, this translates to $\sim 10^{26}$ operations per second to simulate today's population of 7.7 billion humans. It would also require $\sim 10^{36}$ operations over the entirety of humanity's ~ 100 billion lives to date⁸. As we shall soon see, even the latter can be theoretically accomplished with a nano-based computer on Earth running exclusively off its solar irradiance within about one second.

Sensory and tactical information is much less data heavy, and is trivial to simulate in comparison to neuronal processes. The same applies for the environment, which can be procedurally generated upon observation as in many video games. In Greg Egan's *Permutation City*, a sci-fi exploration of simulations, they are designed to be computationally sparse and highly immersive. This makes intuitive sense. There is no need to model the complex thermodynamics of the Earth's interior in their entirety, molecular and lower details need only be "rendered" on observation, and far away stars and galaxies shouldn't require much more than a juiced up version of the Universe Sandbox video game sim.

⁸ Borrowing largely from Bostrom: 10^{11} humans who ever lived * 10^{16} flops * 50 years average life expectancy * 31,556,952 seconds in a year $\approx 10^{36}$ operations needed.

Bostrom doesn't consider the costs of simulating the history of the biosphere. I am not sure that this is justified, since our biological and neurological makeup is itself a result of billions of years of natural selection. Nor is it likely to be a trivial endeavour, even relative to simulating all of human history. Even today, there are about as many ant neurons on this planet as there are human neurons, which suggests that they place a broadly similar load on the system⁹. Consequently, rendering the biosphere may still require one or two more orders of magnitude of computing power than just all humans. Moreover, the human population - and total number of human neurons - was more than three orders of magnitude lower than today before the rise of agriculture, i.e. irrelevant next to the animal world for ~99.9998% of the biosphere's history¹⁰. Simulating the biosphere's evolution may have required as many as 10^{43} operations¹¹.

I am not sure whether 10^{36} or 10^{43} operations is the more important number so far as generating a credible and consistent Earth history is concerned. However, we may consider this general range to be a hard minimal figure on the amount of "boring" computation the simulators are willing to commit to in order in search for a potentially interesting results.

Even simulating a biosphere history is eminently doable for an advanced civilization. A planet-scale computer based on already known nanotechnological designs and powered by a single-layer Matryoshka Brain that cocoons the Sun will generate 10^{42} flops [60]. Assuming the Architect's universe operates within the same set of physical laws, there is enough energy and enough mass to compute such an "Earth history" within 10 seconds - and

⁹ There are $7.7 \cdot 10^9$ humans with $8.6 \cdot 10^9$ neurons = $6.6 \cdot 10^{20}$ total neurons, and 10^{15} - 10^{16} ants with 250,000 neurons $\approx 10^{21}$ total ant neurons.

¹⁰ Though there are good reasons to believe that the total number of neurons on the planet was much lower even 100 million years ago due to the exponential growth in biological complexity over geological time scales [10]. For instance, the humble ant with its 250,000 neurons and relatively advanced cognitive suite - they can pass the mirror test! - evolved from a wasp ancestor 140 million years ago; modern wasps have just 4,600 neurons

¹¹ 10^{26} flops to simulate all of today's humans * ~100 million years (if doublings happened every 50 million years as per [10], we can assume 50% of load happened during last 50M years, 25% around 50M-100M years ago, etc.) * 31,556,952 seconds in a year * ~33.3 (humans currently constitute 3% of Earth's animal biomass, see [59]; assume share of neurons is similar; humans have the highest EQ, but insects benefit from miniaturization) $\approx 10^{43}$ operations needed.

this is assuming they don't use more "exotic" computing technologies (e.g. based on plasma or quantum effects). Even simulating ten billion such Earth histories will "only" take $\sim 3,000$ years - a blink of an eye in cosmic terms. Incidentally, that also happens to be the number of Earth-sized planets orbiting in the habitable zones of Sun-like stars in the Milky Way [61].

So far, so good - assuming that we're more or less in the ballpark on orders of magnitude. But what if we're not? Simulating the human brain may require as much 10^{25} flops, depending on the required granularity, or even as many as 10^{27} flops if quantum effects are important [62,63]. This is still quite doable for a nano-based Matryoshka Brain, though the simulation will approach the speed of our universe as soon as it has to simulate $\sim 10,000$ civilizations of 100 billion humans. However, doing even a single human history now requires 10^{47} operations, or two days of continuous Matryoshka Brain computing, while doing a whole Earth biosphere history requires 10^{54} operations (more than 30,000 years).

This will still be feasible or even trivial in certain circumstances even in our universe. Seth Lloyd calculates a theoretical upper bound of $5 \cdot 10^{50}$ flops for a 1 kg computer [64]. Converting the entirety of the Earth's mass into such a computer would yield $3 \cdot 10^{75}$ flops. That said, should we find that one needs significantly more orders of magnitude than 10^{16} flops to simulate a human brain, we may start to slowly devalue the probability that we are living in a simulation. Conversely, if we are to find clues that simulating a biosphere is much easier than simulating a human noosphere - for instance, if the difficulty of simulating brains increases non-linearly with respect to their numbers of neurons - we may instead have to conclude that it is *more* likely that we live in a simulation.

Computing Costs of Cosmic Expansion

Let us for the time being assume that we need $\sim 10^{16}$ flops to simulate a human brain. This would mean that we would need $\sim 10^{26}$ flops to simulate the current

world population of 7.7 billion (and perhaps 10^{27} - 10^{28} flops to simulate the entire biosphere). This is still many orders of magnitude higher than global computing capacity, which has been estimated to be around $2 \cdot 10^{20}$ - $1.5 \cdot 10^{21}$ as of the end of 2015; assuming median brain simulation requirements (10^{18} flops) and standard rates of growth in computer hardware (25% per annum), these two lines shouldn't intersect until late in the 21st century¹². Under current computing paradigms, this suggests a near absolute guarantee of safety from simulation shutdown to around 2100.

UN projections suggest that the world population will max out at 10-11 billion people by the end of the century. A 2004 meta-analysis of 94 historical estimates of the planet's carrying capacity, which at 7.8 billion is virtually equivalent to today's population [65]. That said, another increase in order of magnitude during the following millennium cannot be excluded if proliferating pro-fertility genes were to reverse the demographic transition and bring the world to a state of "Malthusian industrialism" [66]. Carrying capacity estimates specifically based on land/food as the limiting factor produced a much higher potential population of 33-103 billion, which also syncs with my own estimate of ~100 billion as the planet's carrying capacity under current technological levels [67]. Such a world will require almost 10^{27-28} flops to simulate.

However, these figures will rapidly inflate if/when we reach a "posthuman" stage and start to radically expand our noosphere. This expansion can be either inwards/microscopic (running more and more computations exploiting existing solar potential), outwards/macrosopic (settling nearby star systems, galaxies, supercluster, or the entire universe), or both.

¹² "Global computing capacity" in AI Impacts (2016). <https://aiimpacts.org/global-computing-capacity/>

	Earth	Sun	Earths in Galaxy	Stars in Galaxy	Earths in Universe	Stars in Universe
Only Humans c.2019	1E+10		1E+20		1E+31	
Humans (AoMI)	1E+11		1E+21		1E+32	
"Brains in a Vat" (20W)	8.7E+15	1.93E+25		1.93E+36		1.93E+47
Ems	1E+24	2.2184E+33		2.2184E+44		2.21839E+55
UNITS	1	1	1E+10	1E+11	1E+21	1E+22
ENERGY (joules)	1.74E+17	3.86E+26	1.74E+27	3.86E+37	1.74E+38	3.86E+48

Table 2.1. Population, astronomical, and energy statistics.

Before we take a look at the computational requirements needed to simulate various expansion paths, it would behoove us to first establish some basic numbers. As mentioned above, there are about 10 billion Earth-sized planets within the “Goldilocks zone” of Sun-like planets in our galaxy. There are also approximately 100 billion stars in our galaxy. Our local supercluster contains 50,000 galaxies. There are 100 billion galaxies in the universe. The Sun generates 3.86×10^{26} joules, of which 1.74×10^{17} joules reach Earth. (For comparison, our entire civilization produces just 1.2×10^{13} joules of energy per year¹³). I don’t bother accounting for wastage, since I assume that further technological development will assure that it doesn’t translate into a decline of more than order of magnitude relative to 100% hypothetical 100% efficiency.

There are currently ~10 billion humans on the planet, and I posit that its carrying capacity with today’s technological levels is ~100 billion [67]. Estimates for the human colonization potential of Earth-like planets is a factor of the number of stars and galaxies. I do not think there is much point in trying to estimate the human carrying capacity of like a Ringworld (i.e. in between an Earth-like planet and a Dyson Sphere). If we are to expand to other stars and galaxies in a substantially non-biological form - e.g., as 20 watt “brains in a vat”, as mind emulations, or as superintelligent AI programs - then one may assume that all surface areas will be exploited to maximize the

¹³ This may be radically increased should cheap fusion power be developed. One kilogram of hydrogen to helium fusion, as at the center of the Sun, generates 6.2×10^{14} joules. Consequently, half a ton of fusing hydrogen will generate more power than what the Earth gets in solar input per unit of time. However, there’s no really feasible way to get to the Sun’s output level.

amount of energy tapped from the stars, up to and including the construction of Dyson Spheres.

Robin Hanson argues that sometime during this century, the noosphere will come to be dominated by silicon-based “ems” (emulated minds) [68]. Ems reproduce by copying, and since copying has trivial costs, it can be expected that em society will quickly reach its carrying capacity, depressing wages to subsistence levels. Since much of the brain’s bit erasures are non-logical, Hanson suggests that ems will be much more efficient than the 20 watt human brain and that the Earth will be able to support as many as 10^{24} slow ems¹⁴. This presupposes decades’ worth more progress in increasing energy efficiency in computations (flops per watt), as well as orders of magnitude worth of optimization on brain models (i.e., removing non-logical computations)¹⁵.

Regardless of how much “effective” computing power we manage to squeeze out of ems (or AI software), the ultimate bound on our demand on the Architect’s computing resources is determined by energy and hardware considerations - to which we shall now turn.

	Earth	Sun	Earths in Galaxy	Stars in Galaxy	Earths in Universe	Stars in Universe
Only Humans c.2019	1E+26		1E+36		1E+47	
Humans (AoMI)	1E+28		1E+37		1E+48	
"Brains in a Vat" (20W)	8.7E+31					
Nano-Based Computer	4.5078E+32	1E+42		1E+53		1E+64
<i>Human History (operations)</i>	<i>1.578E+36</i>		<i>1.578E+46</i>		<i>1.578E+57</i>	
<i>Biosphere History (operations)</i>	<i>1.052E+43</i>		<i>1.052E+53</i>		<i>1.052E+64</i>	

¹⁴ Reply to my question on Twitter: <https://twitter.com/robinhanson/status/943112223123230720>

¹⁵ First, the most cost-efficient supercomputer on the Green 500 list as of June 2019 only registers 15 Gflops ($1.5 \cdot 10^{10}$ flops) per watt. The human brain does at least 10^{16} flops per 20 watts, equating to a difference of five orders of magnitude. This metric increased by a factor of ten every decade, so there’s still perhaps half a century left to go, assuming this particular subset of Moore’s Law continues. (Meanwhile, the most powerful supercomputers on the Top 500 list now exceeds 10^{17} flops, perhaps constituting a tenfold superiority over human performance). Second, how are we supposed to know which erasures are “logical” and which are not? After all, the brunt of Hanson’s argument that ems would come before artificial general intelligence rests on the idea that human brains are already here, “ready to go”, and only need to be copied and emulated - as opposed to deeply understood and built up from scratch.

Table 2.2. Computing capacity (flops) needed to simulate various populations of humans, ems, and superintelligences.

One rather surprising consequence of the brain's computational efficiency is that a transition to a neo-Malthusian em or AI civilization running on nano-based hardware will increase the Architect's load by no more than seven orders of magnitude. Since three orders of magnitude and ten orders of magnitude had already been spent on simulating all human history and the biosphere's history, respectively, remaining a Type I civilization on the Kardashev scale is unlikely to put the Architect's supercomputer under undue strain.

However, the equation changes rapidly once we start expanding beyond Earth and its measly share of the solar flux. Every major such expansion - harnessing the energy output of the Sun (Type II), the galaxy (Type III), and the universe - represents an increase of ten orders of magnitude worth of computational potential. Even if expansion was limited to purely biological human colonization of Earth-like planets in the Milky Way, transforming them into 100 billion soul "Hive Worlds" like the Imperium of Man in the Warhammer 40K universe, it will require 10^{37} flops to simulate; that's equivalent to simulating all human history every tenth of a second. Expanding to all potential Earths in the universe would require 10^{48} flops.

The fragility of biological life coupled with the vastness of interstellar distances means that cosmic expansion is likely to be dominated by silicon-based ems or AIs. A typical scenario might involve a von Neumann probe landing on asteroids orbiting a far off star, using the material to construct a Dyson Sphere or Dyson Swarm, and converting the structure into a Matryoshka Brain. Based on prospective nano technologies, simulating just one such structure would require 10^{42} flops. Converting an entire galaxy into Matryoshka Brains would require 10^{53} flops in computing capacity - that's the equivalent of 10^{17} human histories every single second.

It is also possible that posthumanity will invent "computronium" that transcends currently understood limits of engineering; so much so, perhaps,

that “inwards” expansion will long remain more cost effective than “outwards” expansion. In this scenario, it’s feasible that it will eventually approach a one-to-one mapping with the Architect’s hardware assuming analogous laws of physics. Once we approach the Architect’s geographic extent, our simulation will take longer to run than the flow of time in the basement universe.

The Katechon Hypothesis

Forget the power of technology and science, for so much has been forgotten, never to be relearned. Forget the promise of progress and understanding, for in the grim dark future there is only war. There is no peace amongst the stars, only an eternity of carnage and slaughter, and the laughter of thirsting gods. - Warhammer 40K.

If we are indeed in a simulation, there is a risk that the simulation will break down at some stage of these cosmic expansions, forcing the Architect to Ctrl-Alt-Delete us from our sector of space-time. We have no obvious way to quantify at what point this will happen since we do not know how much computing power the Architect have at their disposal, their future time orientation, the priority they allocate to ancestor-simulations, or even whether their universe hews to the same physical laws as ours. As I have argued, the most that we can *weakly* posit is that the Architect is sufficiently invested in us to have run the $\sim 10^{36}$ - 10^{43} operations needed to simulate the evolution of our civilization and/or biosphere, which might not have been especially “interesting” until recently. This establishes a lower bound for the sort of computing resources the Architect has at his disposal.

What effect could this be expected to have on the universe’s geopolitics - its cosmopolitics? Rather paradoxically, the exact value of the simulating supercomputer’s Rmax value - its maximum achieved performance - may not matter nearly as much as the answers to the following questions:

1. Do alien civilizations tend to *believe* that they are in a simulation? Or at least assign a non-trivial chance to the possibility?
2. Once a space-faring civilization spreads beyond the parent solar system, is there any chance of controlling further expansion?

As I shall argue, a certain set of answers here will provide a crisp solution to the Fermi Paradox.

Do Aliens Believe in The Matrix?

Obvious caveat that alien minds are alien, there are a number of good reasons why they might seriously consider the possibility that they're in a simulation.

First, the capacity to imagine the world as illusion seems to have gone together with the evolution of a complex cognitive suite. It has appeared in various forms and throughout diverse cultures in world history, including primordial folk beliefs (the “dreamtime” of Aboriginal Australians), ancient philosophy (e.g. Neoplatonism; Zhuangzi’s butterfly dream), esoteric interpretations of the major monotheistic faiths (Kabbalah Judaism; Gnostic Christianity; Sufi Islam) and heresies (e.g. Cathars; Bogomils). Of these, perhaps the most fascinating “ancient” example is Gnosticism, as it even anticipates the idea of a simulation within a mathematical structure. The existence of the Demiurge - the creator of the material world; a premonition of the Architect? - happens within the scope of the ultimate reality defined by Monad (the One), which is both Bythos (“deepest”) and outside time (“Proarche”), which can be considered a metaphor for the abstract mathematical structures that define the metaverse. So it seems unlikely that intelligent alien beings would have difficulty with the concept.

Second, metaphysics is becoming “digital physics.” Since the publication of “Calculating Space” by Konrad Zuse in 1969, the idea that “computation is existence”, that we live in a “mathematical universe” that can be crisply described as a set of mathematical relations, or that can be modeled as a cellular automaton or computer simulation, has become increasingly popular amongst physicists and philosophers [69–72]. There is the practical observation that reality appears to be discrete at the lowest levels in both space and time (Planck units). The speed of light can be interpreted as the CPU’s clock speed. The universe appears to be extremely “fine-tuned” in a way that is favorable to the emergence of complex life. Furthermore, a great deal of what Einstein called “quantum spookiness” - collapse of the wave function on

observation, or future events determining the past - can be interpreted as the universe making liberal use of simplifying calculations. Just as with Schrödinger's cat, the typical RPG video game doesn't calculate the amount of gold in a treasure chest until you open/observe it. It would be surprising if aliens did not come up with broadly analogous "digital physics" interpretations of reality.

Third, there may appear more telltale signs that we are living in a simulated universe. This could be in the form of what we might call "lazy programming", such as the recent and unexpected discovery that all galaxies rotate at the same speed [73]. Another, closely related set of possible evidence would be cosmic macrostructures that hint at an Architect's involvement. One possible example is a "supervoid" at the CMB cold spot 6-10 billion light years away. It is spherical in shape, ~1,000 times larger than similar voids, and is supposed to be missing ~10,000 galaxies [74]. This translates into 10^{41} flops to simulate that amount of "Hive Worlds", and 10^{57} flops to simulate that amount of Matryoshka Brains (reminder that 10^{43} operations are needed to simulate a biosphere history). Perhaps that region of space "awoke" several billions of years ago, spread in an expanding sphere, and had to be wiped by the Architect? Moreover, perhaps the Supervoid only became so big because it was the first to go into metastasis, and the Architect could wait longer before shutting it down, since computationally intensive biospheres had yet to form in other parts of the universe?

These are admittedly some crazy speculations. That said, it might be possible to devise more grounded tests in the future. For instance, philosopher Michael Johnson argues that apart from ancestor-simulations, there could be two other good reasons to simulate a universe: (1) Instrumentally, to calculate something; (2) Intrinsically, to create a wide variety of interesting qualia [75]. Finding that our universe is optimized for efficient computation, or that all our contingent physical variables are fine-tuned to create maximum positive valence, could potentially greatly increase our confidence in whether or not we reside in a simulation.

Max Tegmark in *The Mathematical Universe* argues that we do not live in a simulation because of the problem of recursion (there is no apparent way to definitively establish you're not in the basement simulation), and because simulating a universe is much more computationally intensive than merely specifying the set of relations between its elements (which is all that his Mathematical Universe Hypothesis requires) [76]. He argues that the existence of a memory stick within our universe containing such a set of relations would not increase the likelihood of finding ourselves in such a universe to any appreciable degree, considering that the Multiverse is infinite in scope anyway. I do not really buy this logic, because even infinities obey the laws of probability. In any sufficiently large portion of spacetime defined by our universe's rules, the chances of us being in a simulation will converge to the quantity of simulated observer-moments divided by the quantity of "real" observer-moments.

Obviously, we can only speculate about those ratios. However, as Bostrom himself points out, the mere fact of us starting up simulations - especially ancestor-simulations - would massively raise the chances that we are within a simulation ourselves, at least so long as we can credibly recreate the observer-moments we experience. That the ultimate reality - the one that the Architect inhabits - may also be purely mathematical has no bearing on whether or not "our" reality is a simulation. The Monad does not rule out a Demiurge. The existence of operating systems does not make impossible virtual machines, nor does it even say anything about the relative ratio of total programs running between the two.

Consequently, the fourth and last point is that if aliens manage to successfully run ancestor-simulations, it will increase their confidence that they are themselves in a simulation. As already mentioned, they will have a variety of reasons to run such simulations. Critically, running simulations isn't likely to strain their computing own computing budget, at least so long as it remains fixed as a percentage of their total computing activity. The catch, of course, is that this may require repeated "prunings" of the simulation.

Finally, it needs to be emphasized that there is no hard requirement that aliens believe or know that they are in a civilization for certain game-theoretic dynamics that will soon be expounded upon to come into play. It is merely sufficient that they *either* (1) assign a non-negligible probability to such a scenario, *and/or* (2) assign a high probability to other civilizations thinking in these terms, *and* assign a greater utility value to the survival and prosperity of their own civilization than to that of other civilizations. In both of these scenarios, even purely utilitarian considerations will dictate a certain set of cosmopolitical imperatives.

Can Expansion Be Controlled?

The biological drive to expand, to exploit more ecological niches, seems to be innate to our species. This is also probably true for most intelligent alien species, since intellect evolves, in significant part, to the challenges of dealing with variegated environments. There is no reason to think space is an exception, as suggested by the sheer plethora of “space opera”-themed films, books, and video games. This sentiment was perhaps most succinctly expressed by Konstantin Tsiolkovsky, one of the founding fathers of rocketry: “A planet is the cradle of the mind, but one cannot live in a cradle forever.”

Once a civilization sets up the requisite economies of scale, there are substantial material benefits to cosmic expansion. Even if increasing a civilization’s power and prestige is too much of an anthropic motivation, there are no end of seemingly more “universal” benefits, such as increasing the energy/computing power at one’s disposal, and providing redundancy against many forms of existential risk. (This is one of Elon Musk’s stated reasons for wanting to set up a Mars base). However, no posthuman civilization, at least in the Milky Way, seems to have made a play for galactic domination. Moreover, we can be relatively sure that nobody in our neighborhood has gone much further than Type II on the Kardashev scale, i.e. fully harnessing the energy of its parent star.

Nor is it immediately obvious why alien civilizations would hold back. After all, it's not like spreading to two, ten, or even 100 new star systems will likely turn out to be the straw that breaks the horse's back (or short-circuits the Architect's supercomputer). Increasing the computational load by many orders of magnitude, such as overspreading a supercluster with ems? As we saw in Part II, this is potentially much riskier. Settling just a few other worlds? Probably not. And the benefits to this are vast, since it would introduce a very substantial safety margin to a civilization's long-term prospects. But all this depends upon the critical assumption that expansion to other star systems can be indefinitely controlled.

The reason is that while "singletons" [55] - which can range the gamut from world dictators to global adherence to a set of ethical norms - are feasible on single planets, they become much more problematic to maintain across multiple planetary systems. The average distance between stars in our part of the Milky Way is around five light years, which puts a massive lower bound on the speed of communications, not to mention physical contact, under currently understood laws of physics. Eventually, one world or another will start to ignore metropolitan edicts against further expansion. This will cause a snowballing effect, since the very fact of expansion will select for more adventurous, rebellious, fecund, and expansive cultures. Furthermore, having already defied the imperial center, this expansive culture will have strong incentives to rapidly maximize its power relative to the coalition that it is now sure to provoke against itself. This, perversely, will make it even more important for it to undertake further rapid expansion.

Historically, one can make the comparison between China and Europe during the Age of Exploration. In the wake of Zheng He's treasure voyages from 1405-1433, the Chinese decided to scrap their navy to focus on the nomadic threat. As a centralized empire, it was able to institute progressive restrictions on private maritime trade, eventually limiting them to tributary missions. China was the dominant Power in East Asia, both culturally and commercially; in neighboring polities, kings sought investiture from and "kowtowed" before the one emperor, the "Son of Heaven". Consequently, the

Chinese sea ban (“haijin”) was also adopted by its cultural vassals, such as Japan (“sakoku”) and Korea (the term “hermit kingdom” predates North Korea). The Japanese policy, which lasted from 1633 to 1853, was even more draconian than China’s, prescribing the death penalty for shipwrecked foreign sailors and Japanese who left the country and then returned, as well as their families and intercessors¹⁶.

Meanwhile, at the opposite end of Eurasia, there was no central authority that could mediate the intensely competitive and expansionist drives of the emerging European nation-states. Even the Pope’s 1494 division of the world between Portuguese and Spanish spheres of influence was sooner a recognition of reality than its creation, and in any case it soon became entirely null and void as other European powers joined in the colonial rush. Moreover, it wasn’t long before even the individual mother countries started to lose control over their settlers. The European colonization of North America became preordained as soon as their settlements became sustainable, despite subsequent efforts by the British to prevent American expansion over the Appalachians. And eventually, almost all of the settler colonies declared independence. Never mind five light-years - even just exercising control over a 5,000 km wide ocean proved too much for Britain, Spain, and Portugal¹⁷.

Furthermore, even controlling a human space civilization is likely to be much easier than policing ems or AI superintelligences. Fundamentally, this is because the latter run on electronic circuits that switch 10 billion times faster than the 20 milliseconds that human brain neurons take to react [78,79]. Robin Hanson projects that the typical em will run at a thousand times human speed, while the very fastest cost-effective ems will run at a million times human speed [68]. One second of thinking for the former will be half an hour for us, while one second of thinking for the latter will be more than ten days for us.

¹⁶ In reality, the sea bans and isolationist policies were far less consistent and draconian than in the popular imagination [77]. But they serve to illustrate the point.

¹⁷ We can find another example, although fictional, in the Warhammer 40K universe. Humanity maintains central control over its galactic imperium through warp travel, which happens at much faster-than-light speed. This is coupled with a fearsome secret police in the face of the Inquisition, and planet-killing weapons that can be unleashed in the event of an “Exterminatus” order. Even so, bureaucratic inefficiency, sabotage, and local discontent still results in thousands of rebellions and defections to Chaos.

In the half hour that an ICBM takes to fly from the US to Russia, a very fast em can live out a typical human life. On a scale of light years, even Alpha Centauri (4.4 ly) will be further away than the Andromeda galaxy (2.5 million ly) so far as very fast em chronology is concerned. Since it is the fastest ems that are expected to have the highest status and influence, this implies that interstellar communications will occur on an em-adjusted chronological scale that's longer than the existence of the human species.

Now assuming that ems and/or AI superintelligence are possible in principle, it seems highly unlikely that any radical cosmic expansion will be based on a biological vector. Not for very long, at any rate. First, as already mentioned, the fragility of biological lifeforms makes prolonged space travel a much more physiologically and psychologically challenging ordeal than for their machine-based counterparts. Second, this probably wouldn't change even if a civilization makes a strategic decision - and has sufficient internal coordination - to ban the development of AI superintelligence (which it might do to reduce the computational load on the simulation, or because doing so would constitute an existential risk, e.g. they work out that friendly AI is impossible in principle, or establish that ems and AIs cannot have conscious experience).

But will a state of "Butlerian Jihad" last over the centuries and millennia, as the number of colonized star systems climbs from the dozens into the thousands and millions? As with the space colonization problem, there need only be one point of failure. Since ems or AI superintelligences on self-replicating spacecraft may be expected to be much faster and more efficient space colonizers than biological ones cocooned within the generation or colony ships of classic sci-fi, they will rapidly overtake and outcompete the latter in the peopling of the cosmos. Moreover, while a transition to electronic-based space colonization may be merely *very likely* in the case of an initial biological expansion, this would rise to a *near certainty* if said initial biological expansion is unauthorized. After all, a planetary subculture that has scant regard for a civilization-wide prohibition on cosmic expansion is

unlikely to take seriously taboos on creating “machines in the likeness of a human mind” either.

In conclusion, it seems that expansion beyond the confines of one solar system vastly increases the chances of further expansion acquiring a metastatic or runaway character due to the practical difficulties of exercising control over distances measured in light years. Conversely, expansionist drives have a good chance of being contained on a single planet. There can be a global treaty mandating planetary isolationism. A sufficiently powerful coalition of countries can subject “rogue” polities that don’t sign up to sabotage, sanctions, or military suppression.

Counterintuitively, this problem may be even *easier* to solve on a *single* em or AI superintelligence planet. This is because any cosmic expansion will be effected through some kind of spaceship, but manufacturing - even at the nano level - would still be much slower relative to the speed of electronic thought, than is conventional manufacturing relative to the speed of human thought. During the time it takes to construct a starship, faster ems will get to experience the equivalent of thousands of years of human thought. During this period, ems that support the isolationist consensus will have plenty of time to discover the project (if it is clandestine), to gather a coalition against it, and to sabotage it. That said, one may also think of counterarguments that suggest cosmic expansion will be harder on an em planet. Hanson posits that em society will be organized around “clans” consisting of multiple copies of the same basic model em. These clans will presumably have very high levels of internal solidarity and coordination. Should a sufficiently large clan take over a planetoid or large asteroid that’s far from major em centers, one may posit that it will be able to construct starships in considerable security underground.

Regardless of whether a human or em civilization can be expected to have a better handle on controlling cosmic expansion, these considerations strongly suggest a hard, natural limit to interstellar expansion under any circumstances: Staying in one’s own solar system.

Space Sakoku and Zero-Sum Cosmopolitics

Summing up, it is perfectly imaginable that advanced alien civilizations may adhere to the following set of beliefs:

1. That they are in a simulation with limited computing power, or at least assign a non-trivial chance to the possibility.
2. That interstellar expansion risks assuming runaway characteristics the more worlds it comes to encompass.
3. That hypothetical alien civilizations come to hold similar beliefs.

These beliefs may lead to a logically deduced and regionally evolved game-theoretical equilibrium that keeps the density of computational activity in any given sector of space low enough that it doesn't overload the Architect's mainframe. This equilibrium may be interpreted as the ultimate metaphysical Katechon, the Biblical concept of "that which withholds" the coming of the End Times in Christian eschatology, and has been variously associated with the Roman Empire, Christian monarchies, and the Hobbesian Leviathan. The Katechon may be expected to manifest itself in a set of behaviors that aims to minimize the probability of runaway cosmic expansion.

What would be the components of such an equilibrium? Imagine we are non-player characters (NPCs) in the Architect's video game. Through recursion, the Architect fears that its own reality is also simulated, and at risk of getting shut down should they exceed the computing power assigned to their sector of space. After all, what would *you* do if an NPC in your video game was to start mining Bitcoin on your computer's GPU for their own benefit? Sooner or later, you'd probably Ctrl-Alt-Delete them. Even more to the point, what would *you* do if you happened to be that NPC?

First, you would refrain from spinning off more NPCs. This means no cosmic expansion, since "child civilizations" will compete for the Architect's limited resources and increase the risk of ruin for everyone. They will live

under the dictums of the heresiarch of Uqbar, a Borgesian character who proclaimed mirrors and copulation to be abominable, as they both multiplied the numbers of men. The solar system would be kept in a state of indefinite sakoku, with an absolute ban on spreading self-replicators, whether biological or machine, beyond it.

Second, you would kill any other NPCs you come across. Shoot first and don't ask later. This consequence of zero-sum cosmopolitics may be termed Mutually Assured Xenocide. It is essentially a modified version of superpredator theory, but with calculating space as the limited resource. Note that there doesn't have to be anything particularly pathological about it, and there might even be pangs of guilt and regret - though perhaps dulled by repetition and existential necessity - as our killers fire up the mirror beam at the epicenter of the radio emissions from our Solar System (provided that they see us first). So it would really be Regretful Mutually Assured Xenocide (RMAX), with Solzhenitsyn's portrayal of life in the Gulag constituting the literary backdrop: "May you die today, so that I may die tomorrow." But fire that mirror beam they will, because it is RMAX that ensures the Rmax delegated to our sector of space is not exceeded.

Third, more speculatively, you would be incentivized to run ancestor-simulations. As already mentioned, posthuman civilizations may tend to do this for a variety of reasons - out of curiosity about their ancestors; to compute something; and/or to investigate a broader range of possible mind-states and qualia (psychonautics). Philosophically, they might also do it to increase their confidence that they are themselves within a simulation by fulfilling one of the conditions within Bostrom's Simulation Argument. If they are "successful" at this, this would make the Katechon Hypothesis much more likely, and would also consequently serve a vital strategic role, such as analyzing how evolved civilizations react to the possibility of themselves being in a simulation (i.e. providing a sample of more than $n=1$), and the cosmopolitical implications thereof (i.e. would this lead to RMAX dynamics?). Ironically, this would also open up an additional incentive for

alien civilizations to fire on sight - as a meta strategy to reduce the risk that they are themselves within a simulation.

The RMAX Equilibrium

Anticipating objections, I need to emphasize that the Katechon would be an *evolved* system. Areas of the universe in which it did not appear, or where RMAX was not enforced rigorously enough, would have been “wiped” by the Architect. On the other hand, the anthropic principle suggests that RMAX was not so rigorously enforced as to have prevented the possibility of life developing on at least some habitable planets. Furthermore, an environment in RMAX equilibrium can also be expected to select for a certain set of psychological traits amongst surviving alien civilizations. These are very likely to include paranoia, isolationism, and aggressiveness.

The fundamental reason has to do with the observation that the offensive rules supreme in space. This is a function of the sheer destructiveness of hypothetical space-based weaponry, as well as the relative ease of stealth for civilizations that are so inclined. Trivially, one may launch dense pellets or objects at very high speeds, because energy imparted is a constant of mass but the square of speed, which furthermore becomes hyperbolic at relativistic speeds¹⁸. However, aiming such a shot may be quite difficult from large distances. Alternatively, mega-mirrors arranged around a star can generate a beam with the intensity of a 6750K blackbody, with a diameter equal to our Moon’s orbit and a diffraction rate of only 50 km per 1,500 light years. Such a beam would maintain the intensity of a star’s surface, even thousands of light years away¹⁹. Even a temporary intersection with a planet such as Earth will fry its surface to a crisp. Another, admittedly harder, possibility is to fling

¹⁸ Fun example from Randall Munroe (xkcd): A 30-meter diamond traveling at 99% of light speed will wreck destruction equivalent to 50 dinosaur-killing Chicxulub impacts (see <https://what-if.xkcd.com/20/>).

¹⁹ See comments by Charles Engelke at “Tabby Star abnormalities in dimming are still consistent with Alien Dyson Swarm construction and long term dimming confirmed with 4 year Kepler data.” <http://www.nextbigfuture.com/2016/08/tabby-star-abnormalities-in-dimming-are.html>

large planets into the Sun²⁰. This isn't going into overly "sci-fi" weapons, such as nanomachines that can transmute a star's mass into elements that don't support fusion (as in Peter F. Hamilton's *The Neutronium Alchemist*), or exotic space-time hijinks to speed up delivery times beyond light speed (e.g. Alcubierre Drives or wormholes)²¹. They might be purely speculative today, but who knows what tomorrow will bring?

Potentially, aliens can also send von Neumann probes programmed to kill, enslave, or otherwise constrain the cosmic expansive potential of native lifeforms. However, this may be risky, since any self-replicators still have the potential to evolve and go "rogue" themselves, nullifying the entire point of such a mission. Perhaps a safer and more productive use for von Neuman probes will be as spying/listening devices on lunar surfaces or asteroid belts, as in Bracewell's sentinel hypothesis [80]. It would be relatively cheap to seed most of the solar systems in a galaxy with such probes, especially those suspected of having habitable planets. In our own solar system, it might make sense to place them within the inner Oort Cloud, which is far enough from the Earth (unlike the Kuiper Belt) to avert detection for what is likely many more centuries to come, but not so far away as to have all radio emissions from Earth fade away into the background cosmic radiation. The low level of solar flux at those distances will deprive self-replicators of the energy surpluses needed for vigorous reproduction and potentially dangerous evolution, but might be just enough to allow them to effect self-repair and passive observation - and to communicate detection of artificially-created radio waves to their masters.

The one thing that all of these attack vectors have in common is that the victimized civilization would hardly have any time to know what hit them, let

²⁰ The kinetic energy released from Jupiter falling into the Sun from rest is equivalent to ~30,000 years of the Sun's output. This would presumably make life in the Earth's current orbit unviable.

²¹ Kinder alien civilizations may instead merely send a heavy object such as a black hole hurtling in our general direction. If it's not accompanied by an accretion disk, we may only notice it pretty late in the game, probably through gravitational microlensing or gravitational effects on the Kuiper Belt. Perhaps this will be just enough time to save the human species before we get ejected out of the solar system by retreating underground and transitioning to geothermal and nuclear energy. Energy surpluses will be very low - geothermal flux is orders of magnitude lower than the solar flux. On a "rogue" planet, our civilization's future potential may be permanently crippled.

alone figure out where it came from. Even in the unlikely event that they regain their bearings, their own detection capabilities would have been massively degraded. Consequently, the civilization that launches the attack would have little fear of retaliation.

Consequently, the correct game-theoretical move under an RMAX Equilibrium is to always defect. Cooperation only typically arises in repeated games - but the distances and scales of space warfare, plus the high risk of attempts at peaceful communication (since they are largely equivalent to dropping stealth), means that there'd be scant space for more positive dynamics to arise. Defection being the rational play would also subdue incentives to actively seek revenge, at least in the unlikely event that a civilization on the receiving end of a space bullet or mirror beam survives in some form.

All this implies that the very nature of the RMAX Equilibrium would actively select for xenocidal aggressiveness. Just as any good or trusting creature dreamt up by mortals and given flesh in the northern Chaos Wastes of the world of Warhammer gets instantly killed by stronger and more evil entities, so too, perhaps, the less paranoid and aggressive space civilizations get snuffed out as soon as they make their existence known to the uncaring gods of the heavens.

One team of futurists has argued that advanced alien civilizations “aestivate”, quietly hoarding their energy surpluses so as to perform computations at a time far in the future when the cooling of the universe makes computing much more efficient [81]. This would enable far more total operations (by a factor of $\sim 10^{30}$) than if they were to start now. They calculate that a civilization burning through the baryonic mass of a supercluster can achieve as many as 10^{93} operations. This should suffice to simulate an entire universe of Matryoshka Brains for up to a sextillion years²². That said, it should be noted that strong counter-arguments have been raised against the Aestivation Hypothesis [82].

²² $\sim 10^{93}$ available operations divided by $\sim 10^{64}$ flops needed to simulate nano-based Matryoshka Brain universe, further divided by the number of seconds in a year = $\sim 10^{21}$ years.

However, even if it is true, it would not constitute a refutation of the Katechon Hypothesis. That is because even aestivating civilizations will need to ensure that upstart civilizations don't emerge and smother them during their slumber, and/or metastasize and invite the Architect's wrath upon their sector of space. Furthermore, the inventors of the Aestivation Hypothesis make the exact same point: "Leaving autonomous systems to monitor the domain and preventing activities that decrease its value would be the rational choice. They would ensure that parts of the originating civilization do not start early but also that invaders or young civilizations are stopped from value-decreasing activities. One plausible example would be banning the launch of self-replicating probes to do large-scale colonization." Consequently, they would if anything have even greater incentives to stymie foreign cosmic expansions than "active" alien civilizations.

More speculatively, combining these considerations may suggest that an optimal strategy under the Katechon Hypothesis may be to enclose a single star within a Matryoshka Brain. The outer shell would constitute a clearly demarcated limit to cosmic expansion. It would give its owners extreme detection capabilities (massive telescopes) and offensive powers (space mirrors). It would also generate up to 10^{42} flops worth of computing power based on theoretically feasible nano-based designs, which could be sufficient to simulate a million human histories within one second. Incidentally, it has been recently theorized that the KIC 8462852 star may be in the final stage of transitioning to a Type II civilization [83]. It is 1,468 light years away from us. If these speculations are correct, by far the strangest thing would be that they have reached a posthuman stage just ~2,000 years ahead of us. Set against cosmic timescales of billions of years, this would be a most amazing coincidence - unless, perhaps, the Architect "seeded" every intelligent alien species at the same point in time in what ultimately translates into a gargantuan Civilization video game. If so, this certainly doesn't bode well for us. We'd have come to a tank battle armed with spears²³.

²³ Let's hope we're playing Civilization 3.

Navigating the Black Seas of Infinity

We have been blithely broadcasting our presence to the dark void above for more than a century. Even if commercial radio or TV broadcasting is too underpowered, the radar signals from Russian and American ballistic missile early warning systems should be detectable from any part of the inner Oort Cloud that happens to host an alien listening post with the detection capabilities of the Arecibo Observatory. If the Katechon Hypothesis is true, then our doom may already be written onto the stars.

Still, it's not yet too late to take some proactive measures to give us at least some chance of survival if worst comes to worst.

(1) We need more research! This sounds banal, but it's true nonetheless. We need to think more about how to prove (or disprove) the Katechon Hypothesis and accurately identify the RMAX Equilibrium's position within the hierarchy of existential risks. In particular, we need to do more of the following:

- Generate much more precise estimates of the likelihood of potential Great Filters in the past. This should be done anyway, since narrowing down the past parameters of the Drake Equation is also critical to clarifying just how much we should be worried about existential risks in principle.
- Continue researching and working to mitigate technogenic existential risks.
- Continue searching for more evidence on whether our universe is a simulation or not.
- Continue working on simulating more complex neuronal structures, to establish the computational cost of simulating minds of varying complexity, and how granular the simulation needs to be to accurately simulate their behavior.

- Explore the nature of qualia, of consciousness, and of whether they can be rigorously measured and simulated.

The answers to these research questions will determine the attention we will need to pay to subsequent recommendations.

Depending on the results of these investigations, it may still be worthwhile trying to enact some form of emissions control, even though there's a good chance it will be politically impossible, and far too late anyway.

In a personal communication, Michael Johnson suggests that we also need to explore what sort of predictions the Katechon Hypothesis implies. For instance, are there cosmological predictions that would add to our confidence about the Katechon Hypothesis if they were later discovered to be true? Possibly the two most likely places to look are cosmological observations and contingent variables in the Standard Model. Does the Katechon Hypothesis make any cosmological predictions, or predictions about what the “apparently contingent but apparently fine-tuned” variables in the Standard Model might be exactly optimized for?

(2) Absolutely no active SETI. One doesn't even need the Katechon Hypothesis to see why this might be a bad idea.

(3) We may consider instituting radio emissions control. This will be politically tricky, though possible for a determined singleton. The main problem is that it's probably far too late for that. Nonetheless, it may still be worthwhile if the deduced likelihood of RMAX Equilibrium is sufficiently and alarmingly high.

(4) We need to get good at identifying small objects in space. If our sector of space is in an RMAX Equilibrium, alien civilizations are likely to have seeded space with secret listening outposts trained to recognize the appearance of intelligent civilizations within their sectors, and relay their findings for all the universe to hear (there's no particular reason that the spotter and the sniper have to be the same civilization).

We will need to comb nearby planetary and asteroid surfaces. As mentioned, there is a good chance that any such Sentinels will be located in the Oort Cloud, exploring which is beyond our present capabilities. However, doing this for nearby planets, moons, co-orbiting objects, and the Kuiper Belt is already on the cusp of technological feasibility.

(5) There must be hard restrictions on interstellar expansion until we can disprove the Katechon Hypothesis. Even if there are no hostile aliens, such an expansion is likely to eventually assume runaway characteristics and doom us to eventual simulation shutdown.

(6) Nonetheless, we need to recognize space technologies as important complements to reducing existential risks. Dispersing our civilization over the solar system would increase the chance that at least some humans would survive the Earth getting fried by a directed energy weapon or hammered by a hypervelocity projectile. Especially prospective avenues might include:

- Early warning systems for incoming projectiles, black holes, mirror beams (if our orbital path is targeted), etc.
- Nuclear pulse propulsion (Orion Drives), by far the most cost effective way to quickly get huge masses of matter out into space.
- Colonization of Mars, Venus, and some lunar bodies, with the ultimate aim of making them self-sustaining.

Although as we have seen there is a strong case for banning cosmic expansion, it may nonetheless be useful to have related technologies on the drawing board should our own planet or solar system come under imminent risk of extermination. This would include life support, life extension, and/or cryopreservation technologies to enable interstellar travel. However, even if we manage to navigate to a habitable planet in another solar system, we will still face the renewed challenge of radio emissions control. This would require research into social technologies or structures that can maintain ideological consistency over the long-term. Alternatively, it may be worthwhile locating geothermally active “rogue” or Steppenwolf planets and brainstorming ways

of colonizing them²⁴. Their location in deep space and lack of a gravitational tether to a star makes them much harder to locate and track, and an underground civilization will have fewer opportunities or need to blast out its presence to the heavens.

(7) We need to be careful about transitioning to a post-biological form of existence. The pros and cons need to be carefully weighed. It is possible that controlling cosmic expansion will be easier for ems or AI superintelligences. On the other hand, merging with the machine would very likely increase the computing load on the Architect's supercomputer by several orders of magnitude²⁵.

(8) If you gaze long into an abyss, the abyss also gazes into you. Should we reach the posthuman stage, we may need to develop our own RMAX enforcement tools - even if it doesn't currently exist within our sector of space. If we conclude with high confidence that the calculating space we inhabit is strongly limited, it would be incumbent upon us to stymie the cosmic expansions of emergent alien civilizations in the future. It is not too early to start thinking about how we might do that as reliably and humanely as possible.

²⁴ "Could we make our home on a rogue planet without a Sun" by Sean Raymond (Aeon): <https://aeon.co/essays/could-we-make-our-home-on-a-rogue-planet-without-a-sun>.

²⁵ My personal intuition is that it's better to stick with our biological hardware - though improved with respect to longevity, intelligence, etc. - so long we cannot be reasonably sure that biological augmentations/optimizations will not result in the loss of consciousness [75]. (Will our noosphere retain any value without conscious beings to experience it?). Besides, it's far from clear that evolution has even come close to exhausting biology's potential for cognitive power. During the coming decades, developments in bioengineering may make pursuit of a "biosingularity" more promising than of Whole Brain Emulation or AI superintelligence. We are not efficiently using the biosphere's existing neuronal stock. A great deal of neuronal activity is locked up in smaller animals, or animals that don't have the morphology to productively exploit it. There is also a great deal of inefficiency within the human species, as suggested by the banal observation that not everyone is a genius. It is doubtful that simulating a Copernicus is significantly more computationally intense than simulating a Cletus. We can "uplift" animals and genetically augment everyone into a Newton or Murasaki Shikibu without seriously cutting into our simulation's computational budget.

Acknowledgements

I am very grateful to Michael Edward Johnson for multiple very helpful and productive discussions, suggestions, and help with editing.

Bibliography

- [1] E. Gaidos, What and Whence II[^]Oumuamua?, arXiv [astro-ph.EP]. (2017). <http://arxiv.org/abs/1712.06721>.
- [2] R. Hanson, The Great Filter - Are We Almost Past It?, (1998). <http://hanson.gmu.edu/greatfilter.html>.
- [3] A. Sandberg, E. Drexler, T. Ord, Dissolving the Fermi Paradox, arXiv [physics.pop-Ph]. (2018). <http://arxiv.org/abs/1806.02404>.
- [4] P.D. Ward, D. Brownlee, Rare Earth: Why Complex Life is Uncommon in the Universe, Springer, 2007.
- [5] N.J. Planavsky, C.T. Reinhard, X. Wang, D. Thomson, P. McGoldrick, R.H. Rainbird, T. Johnson, W.W. Fischer, T.W. Lyons, Earth history. Low mid-Proterozoic atmospheric oxygen levels and the delayed rise of animals, *Science*. 346 (2014) 635–638.
- [6] M.S. Dodd, D. Papineau, T. Grenne, J.F. Slack, M. Rittner, F. Pirajno, J. O’Neil, C.T.S. Little, Evidence for early life in Earth’s oldest hydrothermal vent precipitates, *Nature*. 543 (2017) 60–64.
- [7] M.D. Brasier, J.F. Lindsay, A billion years of environmental stability and the emergence of eukaryotes: new data from northern Australia, *Geology*. 26 (1998) 555–558.
- [8] L.L. Moroz, K.M. Kocot, M.R. Citarella, S. Dosung, T.P. Norekian, I.S. Povolotskaya, A.P. Grigorenko, C. Dailey, E. Berezikov, K.M. Buckley, A. Ptitsyn, D. Reshetov, K. Mukherjee, T.P. Moroz, Y. Bobkova, F. Yu, V.V. Kapitonov, J. Jurka, Y.V. Bobkov, J.J. Swore, D.O. Girardo, A. Fodor, F. Gusev, R. Sanford, R. Bruders, E. Kittler, C.E. Mills, J.P. Rast, R. Derelle, V.V. Solovyev, F.A. Kondrashov, B.J. Swalla, J.V. Sweedler, E.I. Rogaev, K.M. Halanych, A.B. Kohn, The ctenophore genome and the evolutionary origins of neural systems, *Nature*. 510 (2014) 109–114.
- [9] J.M. Nordbotten, N.C. Stenseth, Asymmetric ecological conditions favor Red-Queen type of continued evolution over stasis, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 1847–1852.
- [10] D.A. Russell, Exponential evolution: implications for intelligent extraterrestrial life, *Adv. Space Res.* 3 (1983) 95–103.
- [11] D.A. Skladnev, S.P. Klykov, V.V. Kurakov, Complication of Animal Genomes in the Course of the Evolution Slowed Down after the Cambrian Explosion, *Evolution*. (2013) 249–256.
- [12] S. Mcbrearty, A.S. Brooks, The revolution that wasn’t: a new interpretation of the origin of modern human behavior, *J. Hum. Evol.* 39 (2000) 453–563.
- [13] R. Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, Penguin, 2005.
- [14] P. Turchin, Arise “cliodynamics,” *Nature*. 454 (2008) 34–35.
- [15] A.V. Korotaev, D. Khaltourina, *Introduction to Social Macrodynamics: Secular Cycles and Millennial Trends in Africa*, Editorial URSS, 2006.
- [16] A. Karlin, *Introduction to Apollo’s Ascent*, *The Unz Review*. (2015). <http://www.unz.com/akarlin/intro-apollos-ascent/>.
- [17] C. Murray, *Human accomplishment: The pursuit of excellence in the arts and sciences, 800 BC to 1950*, Harper Collins, 2003.
- [18] W. Roebroeks, P. Villa, On the earliest evidence for habitual use of fire in Europe, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 5209–5214.
- [19] K. Pomeranz, *The Great Divergence: China, Europe, and the Making of the Modern World Economy*, Princeton University Press, 2009.

- [20] E. Buringh, J.L. Van Zanden, Charting the “Rise of the West”: Manuscripts and Printed Books in Europe, A Long-Term Perspective from the Sixth through Eighteenth Centuries, *The Journal of Economic History*. (2009).
<https://www.cambridge.org/core/journals/journal-of-economic-history/article/charting-the-rise-of-the-west-manuscripts-and-printed-books-in-europe-a-long-term-perspective-from-the-sixth-through-eighteenth-centuries/0740F5F9030A706BB7E9FACCD5D975D4>.
- [21] N. Bostrom, Existential risks, *J. Evol. Technol.* 9 (2002) 1–31.
- [22] N. Bostrom, Where Are They?, *MIT Technology Review*. (2008).
<https://www.technologyreview.com/s/409936/where-are-they/> (accessed May 3, 2018).
- [23] K.Z. Stanek, O.Y. Gnedin, J.F. Beacom, A.P. Gould, J.A. Johnson, J.A. Kollmeier, M. Modjaz, M.H. Pinsonneault, R. Pogge, D.H. Weinberg, Protecting Life in the Milky Way: Metals Keep the GRBs Away, *arXiv [astro-Ph]*. (2006). <http://arxiv.org/abs/astro-ph/0604113>.
- [24] G.H. Stokes, D.K. Yeomans, W.F. Bottke, S.R. Chesley, Study to determine the feasibility of extending the search for near-Earth objects to smaller limiting diameters, Report of the Near-Earth. (2003).
- [25] H. Kahn, *On thermonuclear war*, Cambridge Univ Press, 1960.
- [26] C.H. Kearny, *Nuclear war survival skills*, NWS Research Bureau, 1979.
- [27] C. Stager, *Deep Future: The Next 100,000 Years of Life on Earth*, Macmillan, 2011.
- [28] C. Goldblatt, T.D. Robinson, K.J. Zahnle, D. Crisp, Low simulated radiation limit for runaway greenhouse climates, *Nat. Geosci.* 6 (2013) 661–667.
- [29] M. Popp, H. Schmidt, J. Marotzke, Transition to a Moist Greenhouse with CO₂ and solar forcing, *Nat. Commun.* 7 (2016) 10627.
- [30] R.C. Duncan, World Energy Production, Population Growth, and the Road to the Olduvai Gorge, *Popul. Environ.* 22 (2001) 503–522.
- [31] M.A. Woodley of Menie, M.A. Sarraf, M. Peñaherrera-Aguirre, H.B.F. Fernandes, D. Becker, What Caused over a Century of Decline in General Intelligence? Testing Predictions from the Genetic Selection and Neurotoxin Hypotheses, *Evolutionary Psychological Science*. (2018).
doi:10.1007/s40806-017-0131-7.
- [32] M.A. Woodley, J. te Nijenhuis, R. Murphy, Were the Victorians cleverer than us? The decline in general intelligence estimated from a meta-analysis of the slowing of simple reaction time, *Intelligence*. 41 (2013) 843–850.
- [33] A. Kong, M.L. Frigge, G. Thorleifsson, H. Stefansson, A.I. Young, F. Zink, G.A. Jonsdottir, A. Okbay, P. Sulem, G. Masson, D.F. Gudbjartsson, A. Helgason, G. Bjornsdottir, U. Thorsteinsdottir, K. Stefansson, Selection against variants in the genome associated with educational attainment, *Proceedings of the National Academy of Sciences*. (2017).
doi:10.1073/pnas.1612113114.
- [34] M.A.W.O. Menie, H.B.F. Fernandes, A. José Figueredo, G. Meisenberg, By their words ye shall know them: Evidence of genetic selection against general intelligence and concurrent environmental enrichment in vocabulary usage since the mid 19th century, *Front. Psychol.* 6 (2015) 361.
- [35] J. Pietschnig, G. Gittler, A reversal of the Flynn effect for spatial perception in German-speaking countries: Evidence from a cross-temporal IRT-based meta-analysis (1977–2014), *Intelligence*. 53 (2015) 145–153.
- [36] J.R. Flynn, M. Shayer, IQ decline and Piaget: Does the rot start at the top?, *Intelligence*. (2017). doi:10.1016/j.intell.2017.11.010.
- [37] A. Roe, *The Making of a Scientist*, Dodd, Mead, 1953.
- [38] A. Karlin, A. Grigoriev, Модель факторов инновационной эффективности страны, *Siberian Psychology Journal*. 71 (2019) 6–23.

- [39] V. Weiss, *The Population Cycle Drives Human History – from a Eugenic Phase into a Dysgenic Phase and Eventual Collapse*, (2007).
- [40] E. Dutton, Michael A. Woodley of Menie, *At Our Wits' End: Why We're Becoming Less Intelligent and What it Means for the Future*, Andrews UK Limited, 2018.
- [41] M. Kolk, D. Cownden, M. Enquist, Correlations in fertility across generations: can low fertility persist?, *Proc. Biol. Sci.* 281 (2014) 20132561.
- [42] J. Collins, L. Page, The heritability of fertility makes world population stabilization unlikely in the foreseeable future, *Evol. Hum. Behav.* (2018). doi:10.1016/j.evolhumbehav.2018.09.001.
- [43] G. Clark, *A Farewell to Alms: A Brief Economic History of the World*, Princeton University Press, 2008.
- [44] D. de la Croix, E. Schneider, J.L. Weisdorf, Childlessness, Celibacy and Net Fertility in Pre-Industrial England: The Middle-Class Evolutionary Advantage, (2017). <https://papers.ssrn.com/abstract=2896042> (accessed March 3, 2019).
- [45] J.M. Wicherts, D. Borsboom, C.V. Dolan, Evolution, brain size, and the national IQ of peoples around 3000 years B.C, *Pers. Individ. Dif.* 48 (2010/1) 104–106.
- [46] R. Unz, *How Social Darwinism Made Modern China*, *The American Conservative*. (2013).
- [47] J.G. Sotos, *Biotechnology and the lifetime of technical civilizations*, arXiv [physics.pop-Ph]. (2017). <http://arxiv.org/abs/1709.01149>.
- [48] K.E. Drexler, *Engines of creation*, (1990). http://www.academia.edu/download/8464073/engines_of_creation.pdf.
- [49] C. Phoenix, E. Drexler, Safe exponential manufacturing, *Nanotechnology*. 15 (2004) 869.
- [50] V.C. Müller, N. Bostrom, Future Progress in Artificial Intelligence: A Survey of Expert Opinion, in: V.C. Müller (Ed.), *Fundamental Issues of Artificial Intelligence*, Springer International Publishing, 2016: pp. 555–572.
- [51] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, O. Evans, When Will AI Exceed Human Performance? Evidence from AI Experts, arXiv [cs.AI]. (2017). <http://arxiv.org/abs/1705.08807>.
- [52] R. Gruetzemacher, D. Paradice, K.B. Lee, Forecasting Transformative AI: An Expert Survey, arXiv [cs.CY]. (2019). <http://arxiv.org/abs/1901.08579>.
- [53] N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014.
- [54] N. Bostrom, Ethical issues in advanced artificial intelligence, *Science Fiction and Philosophy: From Time Travel to Superintelligence*. (2003) 277–284.
- [55] N. Bostrom, What is a Singleton, *Linguistic and Philosophical Investigations*. 5 (2006) 48–54.
- [56] S. Legg, *Machine Super Intelligence*, 2008.
- [57] S. Pinker, *The better angels of our nature: Why violence has declined*, Penguin Group USA, 2012.
- [58] N. Bostrom, Are We Living in a Computer Simulation?, *Philos. Q.* 53 (2003) 243–255.
- [59] Y.M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth, *Proc. Natl. Acad. Sci. U. S. A.* (2018) 201711842.
- [60] R.J. Bradbury, Others, *Matrioshka brains*, Preprint at [Http://www. Aeiveos. Com/bradbury/MatrioshkaBrains/MatrioshkaBrains. Html](http://www.Aeiveos.Com/bradbury/MatrioshkaBrains/MatrioshkaBrains.html). (2001). <https://www.gwern.net/docs/ai/1999-bradbury-matrioshkabrains.pdf>.
- [61] E.A. Petigura, A.W. Howard, G.W. Marcy, Prevalence of Earth-size planets orbiting Sun-like stars, *Proceedings of the National Academy of Sciences*. 110 (2013) 19273–19278. doi:10.1073/pnas.1319909110.
- [62] N. Bostrom, A. Sandberg, *Whole Brain Emulation: A Roadmap*, Future of Humanity Institute, 2008.
- [63] S. Hameroff, R. Penrose, *Consciousness in the universe: A review of the “Orch OR” theory*,

- Phys. Life Rev. 11 (2014/3) 39–78.
- [64] S. Lloyd, Ultimate physical limits to computation, *Nature*. 406 (2000) 1047–1054.
- [65] J.C.J.M. Van Den Bergh, P. Rietveld, Reconsidering the Limits to World Population: Meta-analysis and Meta-prediction, *Bioscience*. 54 (2004) 195–204.
- [66] A. Karlin, A Short History of the Third Millennium, *The Unz Review*. (2017). <http://www.unz.com/akarlin/short-history-of-3rd-millennium/>.
- [67] A. Karlin, AOMI IV: What Is The Maximum Population Earth Can Support?, *The Unz Review*. (2019). <http://www.unz.com/akarlin/world-population/>.
- [68] R. Hanson, *The Age of Em: Work, Love, and Life when Robots Rule the Earth*, Oxford University Press, 2016.
- [69] K. Zuse, *Rechnender Raum*, 1st ed., Vieweg+Teubner Verlag, 1969.
- [70] S. Wolfram, *A new kind of science*, Wolfram media Champaign, IL, 2002.
- [71] S. Lloyd, Y.J. Ng, Black hole computers, *Sci. Am.* 291 (2004) 52–61.
- [72] M. Tegmark, *The Mathematical Universe*, arXiv [gr-Qc]. (2007). <http://arxiv.org/abs/0704.0646>.
- [73] G.R. Meurer, D. Obreschkow, O. Ivy Wong, Z. Zheng, F.M. Audcent-Ross, D.J. Hanish, Erratum: Cosmic clocks: a tight radius — velocity relationship for H i-selected galaxies, *Monthly Notices of the Royal Astronomical Society*. 477 (2018) 3424–3424. doi:10.1093/mnras/sty792.
- [74] I. Szapudi, A. Kovács, B.R. Granett, Z. Frei, J. Silk, W. Burgett, S. Cole, P.W. Draper, D.J. Farrow, N. Kaiser, E.A. Magnier, N. Metcalfe, J.S. Morgan, P. Price, J. Tonry, R. Wainscoat, Detection of a supervoid aligned with the cold spot of the cosmic microwave background, *Mon. Not. R. Astron. Soc.* 450 (2015) 288–294.
- [75] M.E. Johnson, *Principia Qualia: Blueprint for a New Science*, (2016). <https://opentheory.net/2016/11/principia-qualia/>.
- [76] M. Tegmark, *Our mathematical universe: My quest for the ultimate nature of reality*, Vintage, 2014.
- [77] D.C. Kang, Hierarchy in Asian International Relations: 1300-1900, *Asian Security*. 1 (2005) 53–79.
- [78] M.J. Tovée, Neuronal processing. How fast is the speed of thought?, *Curr. Biol.* 4 (1994) 1125–1127.
- [79] W.R. Deal, Solid-state amplifiers for terahertz electronics, in: 2010 IEEE MTT-S International Microwave Symposium, 2010: pp. 1122–1125.
- [80] R.N. Bracewell, Communications from Superior Galactic Communities, *Nature*. 186 (1960) 670–671.
- [81] A. Sandberg, S. Armstrong, M.M. Cirkovic, That is not dead which can eternal lie: the aestivation hypothesis for resolving Fermi’s paradox, arXiv [physics.pop-Ph]. (2017). <http://arxiv.org/abs/1705.03394>.
- [82] C.H. Bennett, R. Hanson, C. Jess Riedel, Comment on “The aestivation hypothesis for resolving Fermi’s paradox”, arXiv [physics.pop-Ph]. (2019). <http://arxiv.org/abs/1902.06730>.
- [83] T.S. Boyajian, D.M. LaCourse, S.A. Rappaport, D. Fabrycky, D.A. Fischer, D. Gandolfi, G.M. Kennedy, H. Korhonen, M.C. Liu, A. Moor, K. Olah, K. Vida, M.C. Wyatt, W.M.J. Best, J. Brewer, F. Ciesla, B. Csak, H.J. Deeg, T.J. Dupuy, G. Handler, K. Heng, S.B. Howell, S.T. Ishikawa, J. Kovacs, T. Kozakis, L. Kriskovics, J. Lehtinen, C. Lintott, S. Lynn, D. Nespral, S. Nikbakhsh, K. Schawinski, J.R. Schmitt, A.M. Smith, G. Szabo, R. Szabo, J. Vihuo, J. Wang, A. Weiksnar, M. Bosch, J.L. Connors, S. Goodman, G. Green, A.J. Hoekstra, T. Jebson, K.J. Jek, M.R. Omohundro, H.M. Schwengeler, A. Szewczyk, *Planet Hunters X. KIC 8462852 -*

Where's the Flux?, arXiv [astro-ph.SR]. (2015). <http://arxiv.org/abs/1509.03622>.