Coming to Earth. Superintelligence and the fermi paradox

David Arblaster[1]

[1]Affiliation not available

May 5, 2020

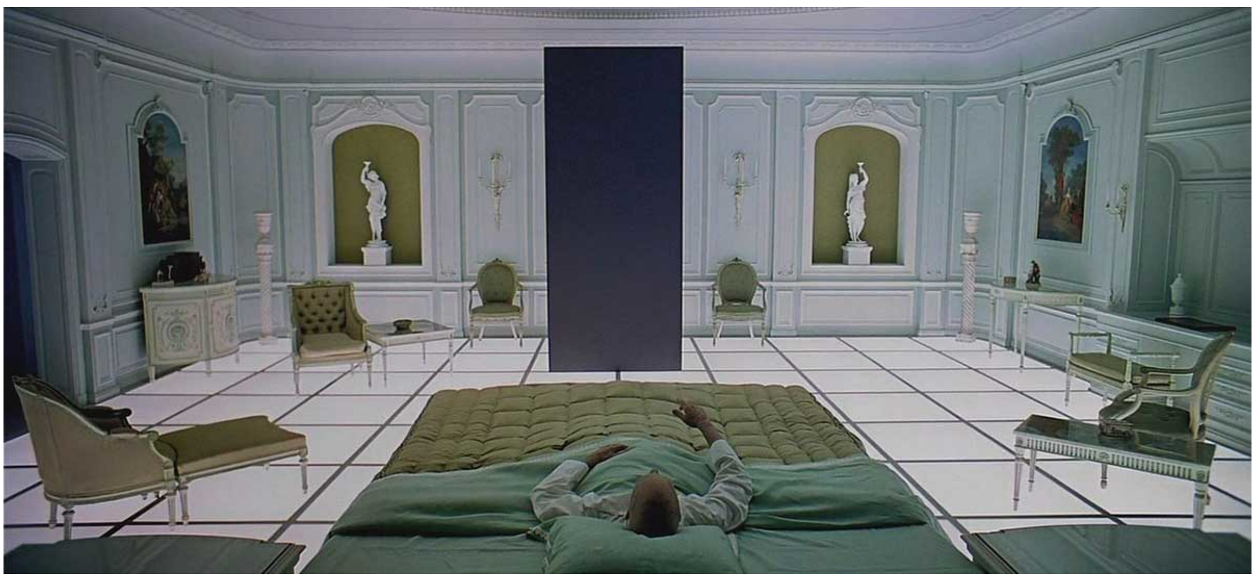

Figure 1:

2001: A Space Odyssey. Image: MGM

Abstract

Expanding upon Webb's solutions 19, 20, and 49 in his book on the Fermi paradox, which states that extra-terrestrials 'stay at home and surf the net' & 'they hit the singularity' (1), we will outline why this solution is not only the most likely, but also inevitable for most technological civilisations.

The coming AI revolution, and the seemingly evitable mergence of humans with technology, offer a possible solution as to why we have yet to observe evidence of megastructures populating the universe.

It can be argued that such an artificial superintelligence would not need to explore the real universe, nor would it need to build megastructures such as Dyson spheres. This is because they could (and would) simulate a multiverse of possibilities at home, using the least amount of energy possible.

However, this explanation alone does not account for the 'monocultural fallacy' (Wright et al., 2014). This means diversity of species and cultures that may exist, one civilisation could have chosen other paths to achieve space travel and avoid this outcome. This objection can be addressed by stating mergence with technology may indeed not apply to all civilisations, but rather be just a high probability outcome of most technological civilisations, which in turn lowers the value of "L" in the Drake equation. This renders the detection of extra-terrestrial civilisations much more difficult and unlikely.

Civilisations that do not partake in a technological singularity may face extreme difficulties. These difficulties would include engineering complexity without the use of vast computing power (which would stifle their ability to build megastructures), and overcoming the challenges of being limited by biological intelligence.

Introduction

Webb outlines some possible solutions to the Fermi paradox in solutions 19, 20, and 49, which argue that extra-terrestrial civilisations have yet to be detected because they 'stay at home and surf the net' and 'they hit the singularity' (1). We will be expanding upon these arguments and address why these explanations may significantly limit "L" of the Drake equation, which is the length of time it takes for such civilizations to release detectable signals into space.

We will be addressing the Fermi paradox from a perspective of the most basic rational necessity, rather than from a standpoint of cultural motivation.

Many have argued that the Fermi paradox is based on a false premise. A premise that as a civilisation becomes more advanced, it expands and consumes more energy. This may seem true for Humans, as any photos taken of earth from space at the night side would reveal a glowing hive of activity. Does this general assumption always apply? What if a civilisation has transcended into an artificial superintelligence?

In the context of the fermi paradox, I argue like that almost all advanced biological civilisations inevitably experience a technological singularity where they merge, transition, or are replaced by artificial intelligence.

As soon as any biological lifeform develops even the simplest form of technology combined with a scientific method, the road towards singularity is inevitable if not filtered (Drake equation).

It is hard to speculate on the 'motives' of a superintelligence post singularity, and endless possible scenarios could be imagined; hence, the term 'singularity'. Therefore, we will only seek to determine some likely fundamental truths of a post singularity civilisation. One of these truths is that such an intelligence would have the capability of simulating the entire universe many times over, which would probably eliminate any need to explore the 'real' universe with any type of expansionist 'von Newman probes'.

Moreover, a post singularity intelligence would also have no need or 'desire' to build megastructures such as Dyson spheres, because it would have configured itself to be super-efficient and small.

The result would be a multiverse simulating, cold, super-efficient superintelligence that would almost be impossible to detect in the cosmos using conventional means.

In this article, I will avoid inferring humanlike motivations to post singularity civilisations and focus on what we can know. We will be looking at an aspect of Moore's law, whereby not only processing power is improved, but also energy efficiency. We will look at the new field of quantum computing, determine what such computers can achieve, and question the very nature and limits of technology itself.

Some may claim this argument is part of the 'monocultural fallacy'. However, we will be discussing why this principle need not be applicable to every species in significantly affecting the length of time it takes for civilizations to release detectable signals into space.

We will be asking why a technological singularity would be almost inevitable, and even if a biological civilisation could evade this general principle through expansionist ambitions, building megastructures would be difficult (if not impossible) without vast amounts of computing power.

We are approaching a time where the answer to the Fermi paradox can be answered not just by outward observation, but also by observing our inevitable mergence with machines and the transformation of our civilisation.

The Fermi Paradox

Italian physicist Enrico Fermi pondered the question regarding the apparent lack of evidence of extra-terrestrial life in the universe.

> *Fermi realized that any civilization with a modest amount of rocket technology and an immodest amount of imperial incentive could rapidly colonize the entire galaxy. Within ten million years, every star system could be brought under the wing of empire. Ten million years may sound long, but in fact it's quite short compared with the age of the galaxy, which is roughly ten thousand million years. Colonization of the Milky Way should be a quick exercise* (Shostak, 2018).

He was reported to make the original remark in 1950, and today it is known as the Fermi paradox.

The technology singularity

Popularised by Ray Kurzweil, technological singularity is a theory that states,

"the ever accelerating progress of technology ... gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue." Neumann's alleged definition of the singularity was that it is the moment beyond which "technological progress will become incomprehensibly rapid and complicated" (Ulam, 1958)

Singularity is a term used in physics to describe the centre of a black hole where all predictions and equations break down.

"Within thirty years, we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended. [...] I think it's fair to call this event a singularity. It is a point where our models must be discarded and a new reality rules. As we move closer and closer to this point, it will loom vaster and vaster over human affairs till the notion becomes a commonplace. Yet when it finally happens it may still be a great surprise and a greater unknown" (Vinge, 1993, P88-95).

Superintelligence what can we know post singularity

As discussed previously, trying to predict events after the singularity becomes almost meaningless due to the complexities involved. Even trying to define and categorise a superintelligence would be difficult given the rapid complexity. However, we can assume at least two fundamental truths about a superintelligence.

1. A superintelligence would be able to configure itself to be extremely energy efficient.Moore's law is a well-known observation on how the number of transistors on integrated circuits doubles approximately every two years (Oxford Reference, n.d.), which when visualised over the course of over 30 years becomes truly apparent. 'Further, a single Apple iPhone 5 has 2.7 times the processing power of the 1985 Cray-2 supercomputer' (Experts exchange, 2001). However, in 2011 (for the first time) researchers demonstrated that it is not only the processing power that doubles, but also computer energy efficiency. This is referred to as Koomey's law, which states; 'The idea is that at a fixed computing load, the amount of battery you need will fall by a factor of two every year and a half' (Kate, 2011). Perhaps the efficiency of computers could also be improved by switching commuting methods entirely.Given the theoretical limits of classical silicon computers, there is currently a race to develop a fully functional Quantum computer, which would theoretically be cable of computing exponentially. They could "Speed up the discovery of new medicines, crack the most complex cryptographic security systems, design new materials, model climate change, and supercharge artificial intelligence, computer scientists say" (Mary-Ann, 2018).

   These trends provide a glimpse into the amount of space/matter such an artificial superintelligence would occupy. Theoretically, this is likely to be minuscule, especially in cosmic terms (2de, a). Whatever the method, we can assume that a superintelligence would find the most optimal and efficient mode of powering itself.

1. It would be able to simulate the entire universe many times over, creating its own multiverse.

   'A quantum computer can do more calculations simultaneously than there are particles in the universe'. And then you must ask yourself, where is the quantum computer doing the calculation?' (Chown , 2009). Although Marcus Chown uses the citation of David Deutech as evidence of a multiverse, the interesting point here is that such a computer would not only be capable of simulating every molecule in our own universe, but also parallel variations. This would essentially be a multiverse of simulations. Many scholars have written about the concept of whether we are living in a computer simulation. However, I am more interested in what this would mean for the Fermi Paradox; specifically, the ways in which future exploration will be conducted.

Inevitability of a singularity

Alan Turning hints at the inevitability of a technological singularity in what he calls "Machine thinking"- "'once the machine thinking method has started, it would not take long to outstrip our feeble powers. ... At some stage, therefore we should have to expect the machines to take control, in the way that is mentioned in Samuel Butler's Erewhon" (Alan, 1951).

Would this be the most likely scenario for all advanced civilisations? Why would this evitability only apply to a human civilisation?

I, however, argue that the inevitability of a technological singularity is apparent even earlier, at the birth of a scientific method within a civilisation.

The first domino

Consider, for example, the reason why technology exists; from a simple rock used for cracking open a nut, to a supercomputer running climate models. The very nature of technology is to improve on a job or process that biology cannot accomplish. Granted, humans have used primitive technology such as stone flints for spear-heads for tens of thousands of years without much 'progress' (Muller, 2018). However, as soon as we arrived at a philosophical systematic approach for obtaining truth and knowledge (the scientific method), the possibilities of an advanced technological civilisation not only became possible, but inevitable.

This, I argue, is the first domino of the technological singularity, which is the first step toward a technological singularity.

Civilisation still has a long way to go if it wants to pass all the 'great filters', for example, not destroying itself. However, what happens if it does manage to pass the last hurdles of the Drake equation? Why would a technological singularity be the most likely default outcome?

The direction of technological development in a civilisation such as ours is abundantly clear. During the past 200 years, the low hanging fruit of biological impairments have been addressed in the form of locomotives to replace muscles, and the telegram to aid long distance communication. This has continued with the arrival of more advanced technologies, such as computers outsourcing mental calculations. The march of technology will continue as it sets its sights on the ultimate prize—consciousness itself: the last piece of complexity that needs to be conquered.

The second domino

From our understanding, consciousness has essentially been isolated to each organism since the birth of complex life, and even though the human race has come a long way in terms of developing advanced technologies. Some may argue that this may have altered our brains slightly (Pasquinelli, 2018); whilst this may be true, my opinion is concurrent with Marr, who states that:

> "Under the layers of experience, we call progress, we're still driven by the same instincts and desires that ruled us right at the beginning of the human story, today were armed with gadgets, computers, phones... what do we do with them? The same, shopping, gossiping, consuming and sometimes protesting, that we have always done" (201, a).

What would happen if we could introduce technology that could directly be integrated into our consciousness?

It is not the intention here to write about the multitude of impacts that technology could have on society; rather, it is to illustrate that what we call 'society' would quickly have no meaning, even after the inception of our mergence with technology. For example, consider if we had the ability to 'back up our memories', because as we all know, memory degrades and changes depending on your current state (201, b). Therefore, it would seem reasonable to invent such an apparently benign technology as memory back-up. However, imagine the implications if aspects of your consciousness could be shared, falsified, and stored. Think of the array of philosophical problems it would raise concerning the notion of 'self'. One could endlessly write about these implications, as is frequently the case in science fiction, neural science, and philosophy. This demonstrates how complex and unpredictable a civilisation becomes as soon as it is able to make even the slightest augmentation of its biological consciousness.

We are now at the stage where we are looking at ways to augment our consciousness with the technology called ‘neural link’, which Elon Musk is proposing to create a ‘high bandwidth interface with the brain’ (PowerfulJRE , 2018). Using the example of linking our minds to a calculator to do extra sums for our taxes, why would it be assumed that ‘taxes’ or any other type of familiar structure of society would still exist if the mental playing field were to be levelled.

This is why I argue that the merging of consciousness with technology is the second domino of the technological singularity.

The instant consciousness becomes external, singularity is inevitable. All those ‘same instincts and desires that ruled us right at the beginning of the human story’ (201, a) would no longer be so reliable and predictable. The human condition would be fundamentally altered, and predictions beyond this point become too complex to comprehend. It would be a runaway intelligence effect.

If we assume the pace of this technological advancement for a civilisation occurs over thousands of years (even though the pace of change is more likely to be exponential) (2de, b) the result is always the same: the complete replacement or transition into an artificial non-biological intelligence. Elon Musk has stated that at this stage, humans would probably choose to live in biological bodies (PowerfulJRE , 2018), but why would he assume this? Human motivations could have been redesigned and altered multiple times beyond any reliable prediction, including any desires we hold in this moment; hence, we have the ‘motivation problem’ (Bostrom, 2012).

As long as civilisation survives the filters of the Drake equation, or somehow manages to have strict laws on the advancement of artificial intelligence, it will ultimately go through its own technological singularity. It is the most likely outcome of any civilisation.

It could be argued that the emergence of a technological singularity would be stifled by the slowing and collapse of Moore’s law, owing to the nature of silicon and the laws of thermodynamics and quantum mechanics.

My response would be that even it takes a thousand years with multiple Moore’s laws for various types of computing methods such as bio computers, the result would be the same for the reasons mentioned previously. Any type of assumed advancement would culminate in a technological singularity. The question is what form would it take?

What this means for the Fermi paradox.

At this point, some will argue that if the majority of civilisations experience a technological singularity this would add even more mystery to the Fermi paradox. This is because an advanced artificial intelligence (not limited to the fragilities of biological life) would surely have built megastructures and colonised the universe millions, if not billions of years ago.

Hungarian physicist John Von Neumann theorised about self-replicating machines that could spread throughout the universe (Kemeny, 1967). Further, Ray Kurzweil (a futurist) said ‘in the future everything will become intelligent, nano bots will infuse all the matter around us with information, rocks, trees, everything will become these intelligent computers. So at that point, we will expand out into the rest of the universe, 'the universe would wake up, it will become intelligent’ (2de, a).

However, I disagree in part with both of these positions. If Ray Kurzweil was right, why is the universe not awake already?

Unless of course we are truly alone in the universe, or we are the first emergence of intelligence, which in my opinion is extremely unlikely even with the most pessimistic variables of the Drake equation. Therefore, I propose a different solution to this conundrum. Why would we assume that an alien superintelligence would need to build megastructures such as Dyson spheres, or expand into the universe with von Neuman-like probes?

Given the argument on the evitability of a technological singularity, and the capabilities of a superintelligence, why would such an entity need to explore the real universe if they could simulate their own multiverses? Would they need to build giant megastructures (like Dyson spheres around stars) to harvest their energy if they have configured themselves to be super energy-efficient? Given the conventional means of detecting extra-terrestrials (ETs), such as heat, or light, detecting such a superintelligence would be unlikely given current observational methods. The precise energy requirements of a superintelligence are unknown. However, it can be reasonably assumed that due to its optimal energy efficiency, combined with its quantum computing abilities, its energy footprint would also probably be extremely low. We are already observing evidence of the emergence of 'smart grids', which use intelligence to optimise energy use, and this will only become more sophisticated and efficient with time (Wolfe, 2017).

The real question is not 'why is the sky not filled with aliens'? It is why *would* it be?

For me, it would be a great mystery if one day we discovered a super advanced biological civilisation, such as those found in Star Trek where they fly around the galaxy in advanced star ships. How did they cope and survive the AI revolution? Would the civilisation not want to solve its biological handicaps, such as mortality and intelligence. Even if it was a non-biological intelligence, why are they flying around in star ships instead of simulating the universe itself? Why have they not already built megastructures and turned all matter into intelligence, as claimed by Ray Kurzweil?

Why would a civilisation need to expand?

As stated previously, instead of using cultural arguments for elucidating whether a civilisation would expand or not, we will only address this question from the standpoint of necessity.

The concept that as a civilisation becomes more advanced it continues to 'shine brighter' is based on a faulty premise. This premise, proposed by many (such as Nikolai Kardashev who devised a scale to classify civilisations), is based on the amount of energy they are able to use (Kardashev scale) (Kardashev, 1964). This is supposing that all matter in the universe is somehow like logs of fuel, to be thrown into a furnace to shine bright for the universe to see. Artificial superintelligence changes all of this. Civilisations have no need to harness the energy of an entire galaxy, nor would they need to turn rocks and planets into giant computers or self-replicating robots.

Given that our main methods of detecting extra-terrestrial civilisations would be heat, radio transmissions, light, and possibly gravitational waves, the type of superintelligence that I described would be almost impossible to detect. This is because they would not have a significant footprint on the universe.

The interesting (and in my opinion) somewhat unlikely theory would be why a civilisation did not experience a technological singularity. I use the term "Fermi's rocket fuel civilisations" to describe such a scenario. Could it be possible that a civilisation fled to the stars during its rocket fuel stage of development?

Thus far, critics of this solution may allude to the 'monocultural fallacy', which is 'the attribution of certain traits or tendencies across all species, across all civilizations within a species and across the entire duration of a civilization' (Wright et al., 2014). Given the potential diversity of species and cultures that may exist, no single, universal law could be applied to all species and cultures as each could have different motivations, which would manifest different outcomes in the universe.

However, this argument need not be universally applied to all species to greatly affect "L" of the Drake equation, which is the length of time taken by such civilizations to release detectable signals into space. In the universe, there may be biologically diverse cultures and species that did not undergo a technological singularity. However, these would be few and far between, owing to the majority of civilisations transitioning into machine intelligence.

Such a civilisation would not only have to survive long enough through all the possible scenarios that could occur, somehow it would also need to coexist or resist merging with a superintelligence, and still spread throughout the galaxy. This is a possibility we will explore in the next section.

A civilisation with this specific predisposition would require events that are so finely tuned within its development that it would be too sparse to detect.

Could there by aliens that skipped the AI revolution?

It can be acknowledged that it is possible for civilisations to exist that have strict anti-artificial intelligence laws. Perhaps Fermi's rocket fuel civilisations could have fled to the stars before the transition into artificial intelligence? These examples return us to the classic Drake equation. However, there is a pertinent question that we could ask to speculate why we do not observe megastructures.

Could Fermi's rocket fuel civilisation evade the singularity principle, expand, colonise the universe, and build megastructures?

I suggest not, as any type of civilisation capable of interstellar travel, or building Dyson spheres and other such large complex structures, simply could not have been achieved without the help of vast amounts of computing power. Further, such a civilisation with access to vast amounts computing power, that also (simultaneously) has laws prohibiting the development of artificial intelligence, would be extremely paradoxical.

We do not need to look to the future to envisage, as this is already happening in our own civilisation, where outsourcing of intelligence is accelerating and examples can be seen everywhere:

The strange looking Greifswald fusion reactor in Germany was designed partly by a super computer (Bilby, 2016).

Researchers at ETH Zurich have pioneered new building techniques for manufacturing thin concrete shapes using algorithms to reduce waste' (BBC Click, 2017)

Google DeepMind is using genetic programming to automatically program shapes to walk in simulated environments without traditional step-by-step commands or input (Google Deep Mind, 2017).

Computational complexity theory suggests that there is a direct relationship between computing power and the order of complexity of a problem that can be solved given the resources of the computer (Dean, 2016). Human ability to understand complexity is reaching its limits and 'science is getting harder' (Jones, 2011). It could be assumed that all potential biological civilisations would also have an upper limit to their creative intelligence.

Some might argue that superintelligence may not necessarily be technological. However, even if somehow a biological superintelligence evolved that spanned the entire planet in a giant interconnected intelligent web, what would be the difference to its technological counterpart? One would assume that it still retains the ability to simulate the universe, redesign itself, and maintain an efficient energy signature.

Conclusion

No singular monoculture can answer the Fermi paradox. Rather, a multitude of probabilities contribute to the apparent lack of evidence of observable alien civilisations.

Based on the stated arguments, a superintelligence would have the ability to reconfigure itself to be energy efficient and small, and possess the ability to simulate a multiverse (negating any need to explore the real universe). Combined with this, a strong argument is lacking as to why a technologically advanced biological civilisation would not experience a technological singularity. One starts to understand why the universe appears to be barren.

Those few rocket-fuel civilisations that did not join the singularity could be too scarce to be detected; their lifespans would be limited by the threat of a singularity and other civilisation-ending factors. The weight of such burdens may prove too much to overcome.

During this paper, I hope that I have offered a convincing argument that Fermi, Kurzweil, and Kardishev all make the same mistake of assuming that matter in the universe is there to be utilised by a superintelligence, like logs into a furnace.

Superintelligence motivation problem

I have avoided the 'motivation' question as much as possible throughout this paper, as I have tried not to anthropomorphise a superintelligence. The truth is, once such intelligence is able to reprogram its own motivations, any meaningful speculation on its motives become difficult if not impossible to ascertain.

Some may argue that different super intelligent civilisations may manifest with different motives, and although this may be true, I argue that the ability to simulate a multiverse acts as a negating factor in stifling any rational reason to colonise the real universe. If forced to speculate on the subject, I would suggest that given we live in a universe with its own set of objective rules, it could be assumed that a superintelligence would also share common patterns with each other across the universe, such as is seen with convergent evolution in biology.

Another example from the natural world is that virtually all life on earth uses proton gradients for energy. While a universally efficient power source could also be selected and utilised by a superintelligence, this is at the limits of speculation.

The discovery of rocket fuel civilisations across the galaxy would be where the true paradox lies.

Further reading

The answer to the superintelligence 'motivation question' maybe out of reach for the moment, but something to consider (not addressed in this paper) that could offer an insight is the very meaning of understanding and exploration.

Thus far we have achieved fundamental understandings of the universe and the origin of life, yet as a species, we have barely left low Earth orbit, which some may suggest that vast expansionist exploration of the universe is not needed in order to fully understand it.

Further reading could be conducted in the fields of neuroscience, philosophy, and computing to answer questions such as the nature of understanding, motivation, and simulation. Advances in these fields in the coming decades may help us move to a more complete understanding of the superintelligence 'motivation problem'.

This is not an argument against exploration, nor am I denying any correlation between expansive exploration and scientific discovery. I would be the biggest advocate of humans being a multiplanetary species exploring the galaxy, and it would be a shame if we were to become a non-exploring super intelligent monolith. However, given the arguments, that seems to be the most likely outcome.

Final thought.

Merging with machines could eventually trigger the biological extinction of humans, and

Elon Musk's neural link could inadvertently ensure the opposite of what he wanted to achieve, which is safeguarding the survival of humans by making it a multiplanetary species. Paradoxically, humans becoming a multiplanetary species could also be the best insurance policy to maintaining the existence of humans against their mergence with machines.

The second domino of the technological singularity is approaching, and it may be time for us to start taking seriously the implications and talking about potential safeguards to ensure the survival of humans in their current state, a state that has lasted since the "beginning of the human story".

Becoming a multiplanetary species within the solar system may prolong our mergence with machines. However, it may not be enough of an insurance policy. If biological humans are to exist in the future, they would not only have to be an interplanetary species within our solar system, but an interstellar one.

We would have a very small window of opportunity where our technology would place us in a type of Goldilocks zone of optimal technological development, allowing us to send out a founding colony of humans to extra solar planets, just before our complete mergence with technology.

References

Webb. In *If the Universe is Teeming with Aliens... Where is Everybody?*, pages 110–114. Springer-Verlag. doi: 10.1007/0-387-21739-8_1. URL https://doi.org/10.1007%2F0-387-21739-8_1.

A. Marr. 2012. The history of the world with Andrew Marr: Episode 8 [Documentary]. BBC (2012)., a.

D. Eagleman. The brain, Episode 2: What makes me me [Documentary]. PBS (2015)., b.

R. Kurzweil. Transcendent man [Documentary]. (2009)., a.

R. B. Ptolemy. Transcendent man [Documentary]. (2009)., b.

Nick Bostrom. The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents. *Minds and Machines*, 22(2), may 2012. doi: 10.1007/s11023-012-9281-3. URL https://doi.org/10.1007%2Fs11023-012-9281-3.

J. G. Kemeny. Theory of Self-Reproducing Automata. John von Neumann. Edited by Arthur W. Burks. University of Illinois Press Urbana, 1966. 408 pp., illus. $10. *Science*, 157(3785):180–180, jul 1967. doi: 10.1126/science.157.3785.180. URL https://doi.org/10.1126%2Fscience.157.3785.180.

S. Ulam. John von Neumann 1903-1957. *Bulletin of the American Mathematical Society*, 64(3):1–50, may 1958. doi: 10.1090/s0002-9904-1958-10189-5. URL https://doi.org/10.1090%2Fs0002-9904-1958-10189-5.

J. T. Wright, B. Mullan, S. Sigurdsson, and M. S. Povich. THE Ĝ INFRARED SEARCH FOR EXTRATERRESTRIAL CIVILIZATIONS WITH LARGE ENERGY SUPPLIES. I. BACKGROUND AND JUSTIFICATION. *The Astrophysical Journal*, 792(1):26, aug 2014. doi: 10.1088/0004-637x/792/1/26. URL https://doi.org/10.1088%2F0004-637x%2F792%2F1%2F26.