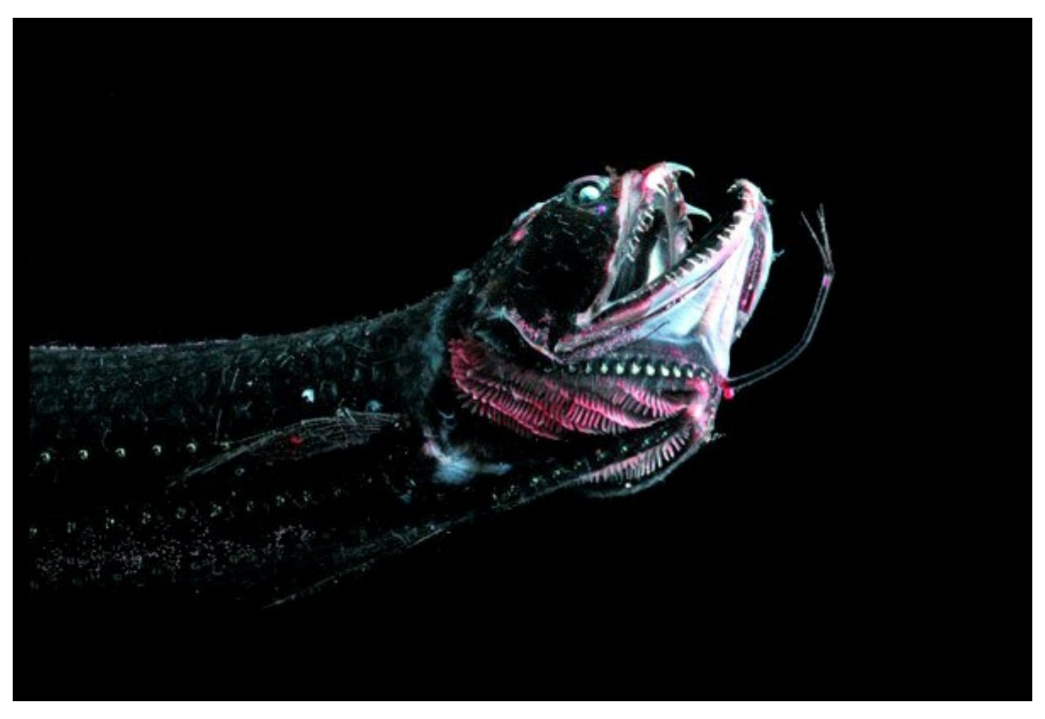
Hunters in the dark: game theory analysis of the deadly probes scenario

Anders Sandberg and Stuart Armstrong

Future of Humanity Institute, Oxford University

Introduction

One group of proposed explanations for the Fermi paradox is the ”deadly probes scenario”, where some civilizations produce self-replicating devices that prevent other civilizations from coming into being [2]. Whether this kind of scenario works as an explanation depends on whether it is stable and compatible with our own observations (including our own existence as observers). This paper analyses these conditions in more detail.



Sharpening the Fermi question In previous work we have shown that given certain modest assumptions about automation and technological feasibility intergalactic colonisation using automated replicating devices is possible using local resources and high fan-out [1]. This model enlarges the volume in which alien civilizations could have been able to reach us between 8 and 9 orders of magnitude compared to earlier galactic colonization models, straining many proposed explanations (rarity of life, high existential risk, strong cultural convergence). However, our model is very compatible with the “zoo” hypothesis or the deadly probe scenario, since it gives early species the chance to deploy probes implementing their policies in a vast volume. Once emplaced probes can remain essentially indefinitely [3].

Our approach

In order to work as a Fermi explanation the replicators must fulfil four conditions:

1. **Cause the great silence.**
2. **Be compatible with our existence.**
3. **Be silent enough in their activities not to be visible.**
4. **Be impossible to overthrow by an new civilization.**

This leads to two questions: *Q1: Is 3 compatible with 4? Q2: Is 2 compatible with 1 in the deadly probes scenario?*

Question 1: Stability

There are two types of stability in the scenario. **Type 1 stability** consists of a dominant species of probe can prevent a local emergent species from producing and launching probes. **Type 2 stability** means coexistence of two or more probe species, with neither able to get upper hand.

Type 1 stability

If a probe becomes visible, the probability that there is no enemy probe within radius r is $\exp(-4\pi r^3\rho/3)$, where ρ is the density. The emergent species will try to build and launch N probes. They can be detected within radius r_1 during the building, r_2 when launching and r_3 when the probes deaccelerate at their destinations. We assume deep space coasting is practically invisible. The expected number of surviving daughter probes will be $N\exp(-4\pi(r_1^3+r_2^3+r_3^3)\rho/3)$. If this is >1 then the species cannot be inhibited. Hence

$$\rho > (3/4\pi)\log(N)/(r_1^3+r_2^3+r_3^3)$$

is a stability criterion for the ”zoo” hypothesis and effective deadly probes.

If $r_i \approx 50$AU, $N = 4\cdot 10^9$ [a], $\rho > 10^{-5}$ per AU, or about 7.4 per solar system. With a one parsec detection range there is a need for 1.8 per pc^3.

This analysis assumes one enemy probe is enough to inhibit a launch of N probes. If the launch step instead requires N inhibitors the density needs to scale as $\approx\sqrt{N}$ per detection volume, requiring $\approx 63,000$ probes per solar system.

Conclusion: For condition 4 to be compatible with 3 probes need to be very stealthy or have a long range of operations.

Type 2 stability

We investigate type 2 stability by an agent-based model. Emissions from attacks, accelerating or decelerating probes are detectable within a fixed radius, while dormant or coasting probes are undetectable. Different probe ”species” are assumed to have equivalent technology. Attacks are local: a probe needs to reach a location in order to damage something.

Consider a model of randomly distributed probes of two species in 3-space with average density ρ_1 and ρ_2. At $t=0$ one probe of species 1 becomes visible due to an acceleration or reproduction event, alerting species 2 probes within a radius. They have a choice to attack, in turn becoming visible to species 1 probes, and so on. Each attacking probe will move towards its target but may be intercepted if an enemy probe targeting it is closer. The whole chain of attacks is resolved in time order; the goal is to estimate the probability that the initial probe survives. We assume probes cannot retask if their target is destroyed.

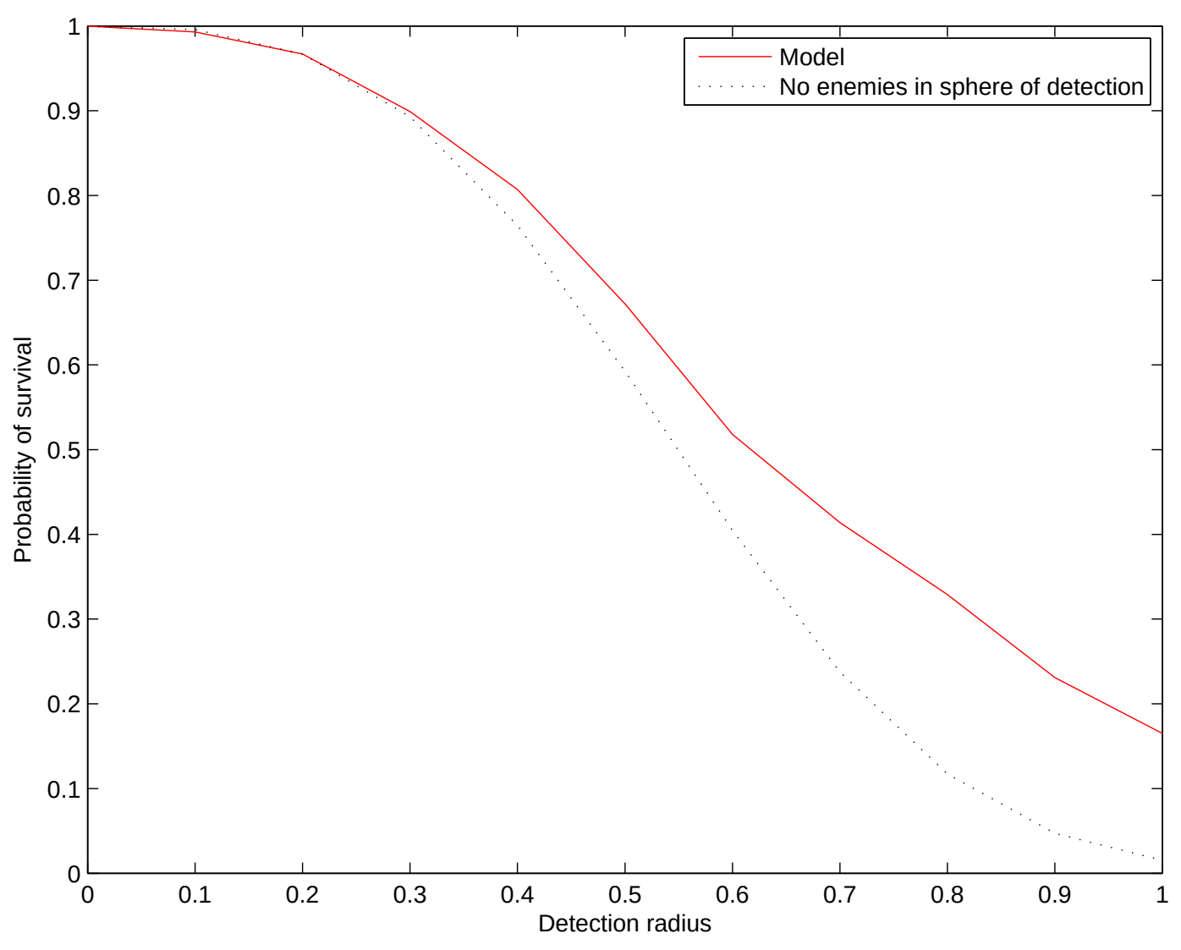


Figure 2: Probability of survival for visible probe as function of detection distance. Area above dotted curve represents where it was “saved” by friendly probes from attack. $\rho_1 = \rho_2 = 1$.

Even for very low densities of friendly probes a probe has a decent chance of escape as long as the enemy density is not large enough to reliably cover the detectable volume (similar to type 1 stability). Hence, dominant species must manufacture enough probes to cover space thoroughly in order to prevent non-dominant species from bootstrapping.

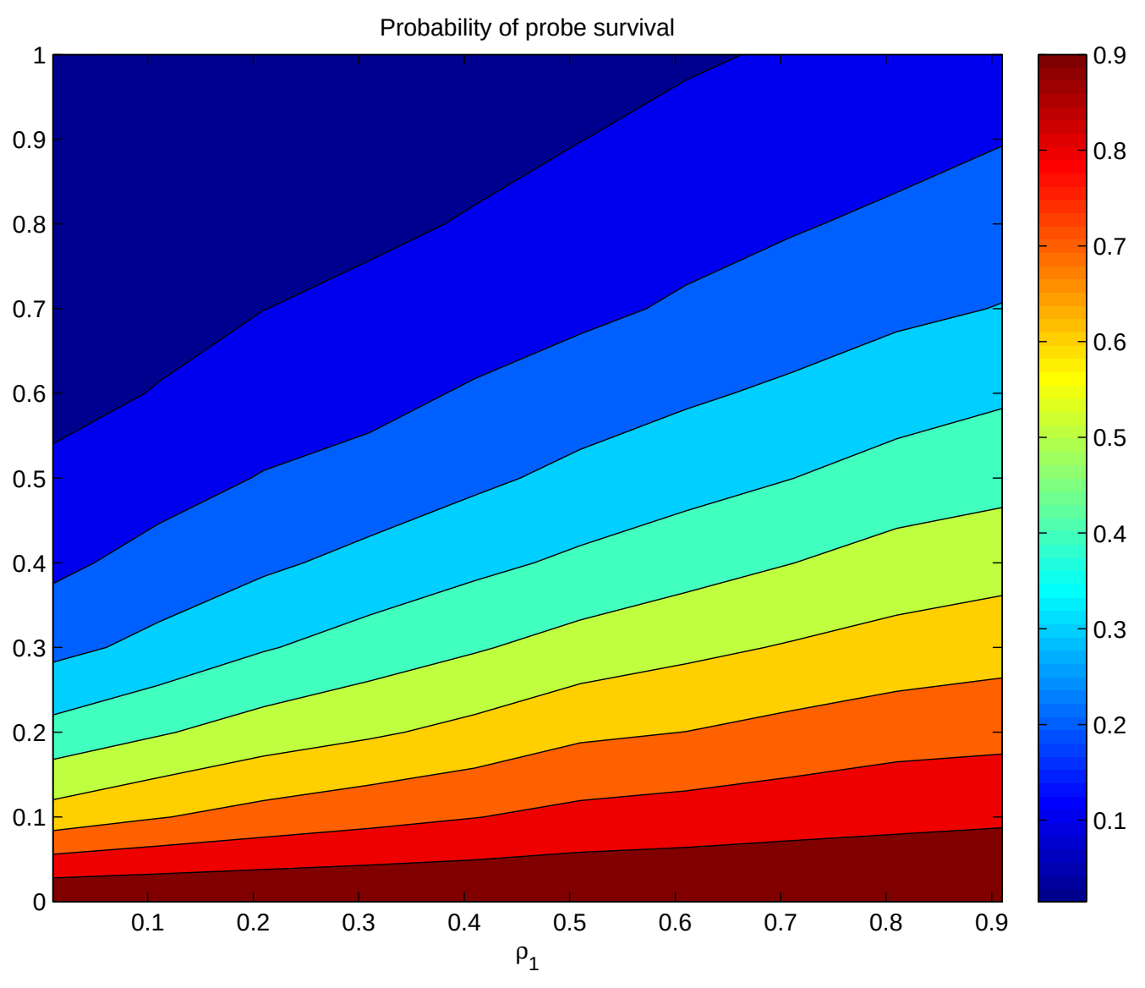


Figure 3: Probability of survival for visible probe as function of probe densities. Detection distance=1.

For higher densities this case is similar to the Lanchester ”linear law” of attrition warfare since losses are approximately proportional to the product of the densities, giving a linear advantage to the the most numerous side.

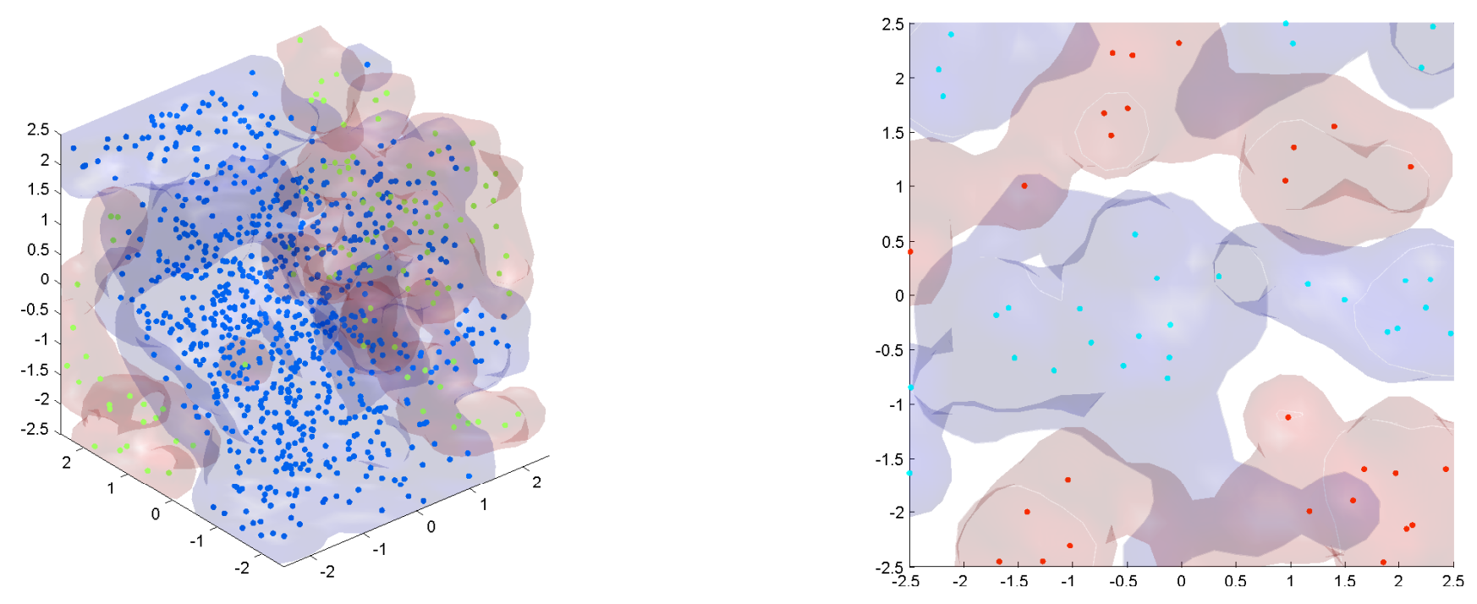


Figure 4: If probes reproduce the typical result is spatial segregation. Regions of width ≈ the detection radius are dominated by one species, separated by empty ”no man’s land”. Reproduction inside the regions will be unhindered, and a gradual coarsening of the boundaries occur.

Conclusion: Both the high and low density cases favour each side manufacturing more probes, causing more attacks and resource use. Since the only upper limit is availability of resources the equilibrium state corresponds to using all available resources. Condition 3 will hence be broken.

Question 2: Anthropic limitations to deadly probe scenarios

An obvious argument against deadly probes is our existence. However, this only works as an argument against *perfect* deadly probes: if the probes accidentally leave some civilizations to develop, even with very low probability, then observers will be rare but observe an apparently probe-free environment.

Not very good deadly probes: Ruled out by condition 1.

Unusual location: This is ruled out by the typical galactic disk location, low-inclination orbit and temporal position of the sun. Had sol been a halo star or extremely early this explanation would have worked.

Not got to us yet: Probes are not omnipresent, but will move in to quench emerging civilizations shortly after they reach some detectable state of development. Pre-empting civilizations requires reliable attack long before there is any chance of launching probe technology. The longest travel time compatible with causing the great silence must be significantly shorter than shortest time from detectability to launch for *any* civilization[a].

Toy model A toy model demonstrates the problem of maintaining pre-emption: Assume civilizations would take a lognormally distributed time from becoming detectable to able to achieve launch. We set the median e^μ to 200 years, with a scale parameter $\sigma = 0.25$ corresponding to a variance of $\approx 50^2$ years. If there are N civilizations, the probability that at least one will escape destruction if it occurs at time T is

$$P(\text{escape}) = 1 - (P(\text{no escape}))^N$$

$$= 1 - (1 - F(T))^N = 1 - \left(\frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{\log(T)-\mu}{\sqrt{2\sigma^2}}\right)\right)^N$$

As can be seen in figure 5, although this distribution has very few civilizations expand before 100 years, even for modest N the required time until destruction has to be a few decades. For wider variance attack has to be even faster.

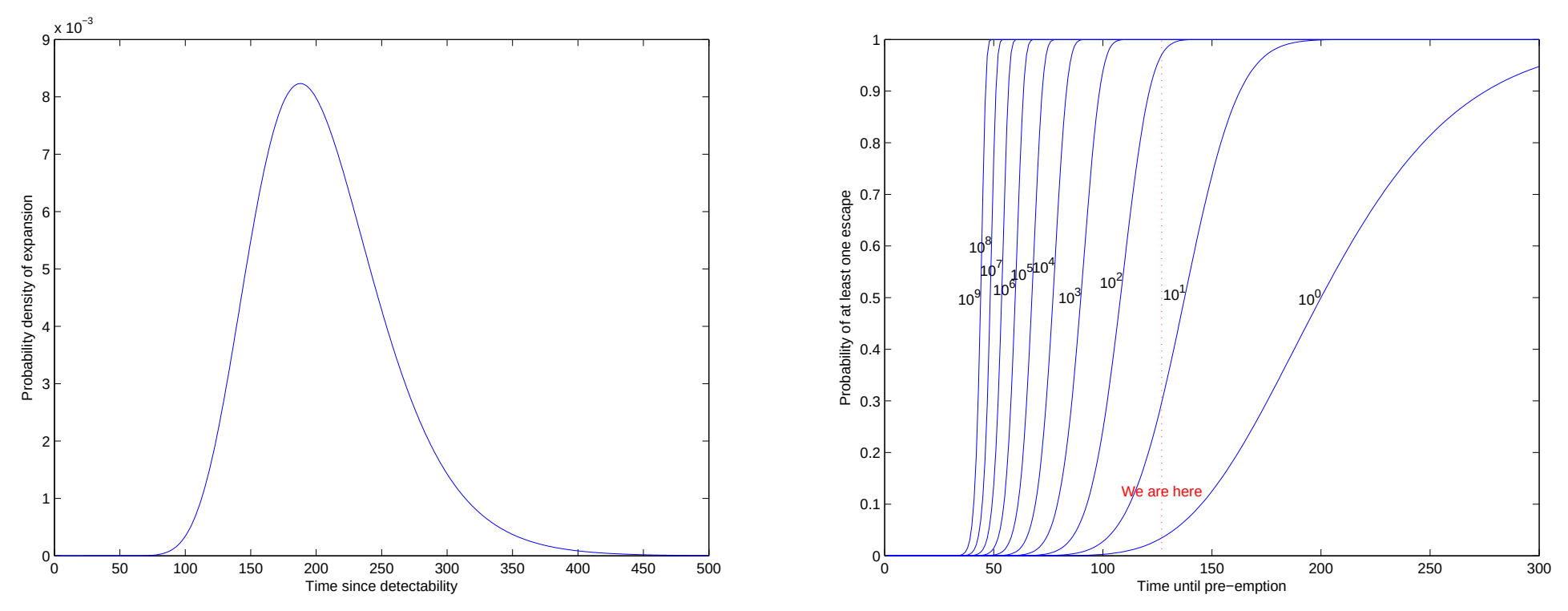


Figure 5: Toy model:(left) Distribution of escape times (right) Probability of unsuccessful containment for different number of civilizations.

Conclusion: The delayed attack scenario is unlikely given our continued existence, unless combined with a rare-intelligence explanation.

Conclusion

It is hard to prevent breakout of a technological civilization capable of building replicators, inefficient or delayed pre-emption can allow civilization to achieve this capacity, efficient pre-emption should have destroyed us by now, so: since we still exist, we conclude that deadly probes are not the main cause of the Fermi paradox.

Acknowledgements We are grateful to Milan Cirkovic, Carl Shulman, Robin Hanson and Eric Drexler for ideas and comments.

References

[1] Stuart Armstrong and Anders Sandberg. Eternity in six hours: Intergalactic spreading of intelligent life and sharpening the fermi paradox. *Acta Astronautica*, 89(0):1 – 13, 2013.

[2] G. D. Brin. The Great Silence - the Controversy Concerning Extraterrestrial Intelligent Life. *Quarterly Journal of the Royal Astronomical Society*, 24:283–309, September 1983.

[3] Anders Sandberg and Stuart Armstrong. Indefinite survival through backup copies. Technical Report 2012-1, Future of Humanity Institute, Oxford University, 2012.

---

[a] The number of probes in the upper-end case in [1].

[a] In the case of humanity, we might have become detectable after 1886 due to radio emissions, and could *at least hypothetically* have achieved large-scale space activity 150 years later.

Address for correspondence: `anders.sandberg@philosophy.ox.ac.uk` Future of Humanity Institute, University of Oxford, Suite 8, Littlegate House, 16/17 St. Ebbe’s Street, Oxford, OX1 1PT