DOCUMENT

UFD020 Milan Ćirković

Close Encounters of the Philosophical Kind

In a discussion of the epistemological questions that are central to the search for extraterrestrial intelligence, Milan Ćirković insists that philosophical thinking is an indispensable part of astrobiology and SETI research.

> MONSIEUR JOURDAIN: There is nothing but prose or verse?
> PHILOSOPHY MASTER: No, Sir, everything that is not prose is verse, and everything that is not verse is prose.
> MONSIEUR JOURDAIN: And when one speaks, what is that then?
> PHILOSOPHY MASTER: Prose.
> MONSIEUR JOURDAIN: What! When I say, 'Nicole, bring me my slippers, and give me my nightcap,' that's prose?
> PHILOSOPHY MASTER: Yes, Sir.
> MONSIEUR JOURDAIN: By my faith! For more than forty years I have been speaking prose without knowing anything about it, and I am much obliged to you for having taught me that.
> — Molière, *Le Bourgeois Gentilhomme*

Monsieur Jourdain's astonishment should be called to mind whenever a scientific (or religious or political, etc.) vigilante tries to put down or sneers at philosophical analysis of any problem under discussion. In contrast to the conventional—and wrong—understanding of this scene, while speaking in prose is in itself trivial, the insight itself is highly nontrivial, especially in the context of education and outreach—as is the insight that intelligent beings do many things without being aware of their place in sophisticated classification schemes. As it happens, vigilantes are in the right sometimes, and some philosophical analyses are tantamount to inventing fancy words for well-known concepts and clearly justified procedures. More often, though, especially when 'big issues' are concerned, vigilantes manage to sound just as ridiculous as Monsieur Jourdain, by failing to position their thoughts and actions in a wider scheme of things. After all, Jourdain needs help with writing a love letter—something clearly of high importance and complexity in his life and social aspirations.

The 'big issues' we will deal with here are of a rather different order: problems such as the origin of life and intelligence, the universe or multiverse, and Fermi's paradox—the apparent contradiction between the high probability of other intelligent civilizations apart from ours, and the lack of evidence for them. And our point here is that, if the philosophical questions and presuppositions involved in addressing these issues are not acknowledged, we may well, like M. Jourdain, be in the position of ourselves 'speaking philosophy without knowing it'.

A philosophical perspective is not an option as far as young fields such as astrobiology and SETI are concerned; it is unavoidable

A philosophical perspective is not an option as far as young fields such as astrobiology and SETI are concerned; it is unavoidable. From the very beginning, the search for life elsewhere has had an important epistemological content: the problem of recognizing life which evolved in an environment sufficiently different to the terrestrial one is obviously something which cannot be resolved on a purely empirical level, with our present and near-future insight into fundamental biological processes. The same applies to more complex questions about general astrobiological complexity and its evolution; and to an even greater degree to the concerns encountered specifically in SETI studies.

One part of philosophical baggage that we should leave at the entrance, though, is the misleading insistence on definitional issues. To precisely define either life or intelligence is impossible at present, as readily admitted by almost all biologists and cognitive scientists. This, however, hardly prevents any of them in their daily research activities or even in synthetic thinking about the cutting-edge problems in their fields. They intuitively understand what the history of science thus far has amply demonstrated, namely that formalizations and definitions come *after* most of the research in a field is done, and not before. The deeply misguided idea of 'definitions first!' is a relic of the epoch of logical positivism and its explicit or implicit faith in inductions and formalizations; such an approach characterizes pre-Gödelian and pre-Popperian thinking about science and truth. The more universal and ubiquitous a concept is, the tougher is the definitional challenge.

The real issue is to what extent we can achieve an intuitive grasp of problems in the field, which is something very different from having a formal definition

If anything, the 'exotic' nature of astrobiology and SETI will make caution in formalization even more pronounced. The real issue is to what extent we can achieve *an intuitive grasp of problems in the field*, which is, of course, a prerequisite for successful research, and is something very different from having a formal definition. By the same token, the lack of formal definitions will not impede or obstruct research in these fields, just as the lack of a formal definition of life did not impede or obstruct research in evolutionary biology, and the definitional ambiguity about the definition of the universe did not impede or obstruct research in cosmology. The powerful and widely present desire to castigate SETI as 'definition-challenged', therefore, tells us more about deeper and irrational animosities toward SETI than about the enterprise itself. Young fields are, almost by default, lacking in formalism, clear definitions, and other forms of orderliness. Such was physics in the time of Galileo, chemistry in the time of Lavoisier, psychology in the time of Wundt. How could it be otherwise? And yet for multiple reasons—some of which have to do with cozy anthropocentrism permeating most of human culture—a similar degree of slack is not cut for SETI. But, just as the intuitive concept of life enables research into terrestrial biology and other life sciences, intuitive recognition of cultural properties such as advanced technology, willingness to communicate, xenophobia/xenophilia, etc. are necessary for formulating SETI activities, although, like the scaffolding, they may be discarded at a later stage and in face of better understanding.

Such an approach goes hand-in-hand with the assumption that, in this young field, there are still philosophical issues at stake, issues that we should not try to hastily 'define' out of existence. Below I shall consider four all-important philosophical assumptions which are (usually tacitly or even unconsciously) built into all formulations of Fermi's paradox.

Philosophical Naturalism

Methodological naturalism tells us that we should not invoke supernatural agencies and capacities in searching for explanation of observed phenomena. This clearly does not presume any attitude toward the existence of such supernatural agencies, only that our explanatory mechanisms do not invoke them either explicitly or implicitly. Emerging from Enlightenment secular thought, methodological naturalism is best encapsulated in Laplace's famous sentence directed to the First Consul of the French Republic, Napoleon Bonaparte: *Je n'avais pas besoin de cette hypothèse-là*.[1]

1. 'I had no need of that hypothesis.' It is important to understand that Laplace did not object to the existence of God per se, but instead rejected Newton's hypothesis that only occasional Divine intervention keeps orbits in the Solar System continuously stable; see also Stephen Hawking's comments at <http://web.archive.org/web/20000708041816/http://www.hawking.org.uk/lectures/dice.html>.

Obviously, methodological naturalism is the basis of all science, the very bread and butter of both everyday research and our grand theoretical syntheses. It has been immensely successful and has not encountered any significant obstacle so far. While this in itself does not imply that it will continue to be successful in the future—that would be a bad form of inductivism—we certainly do not have any reason to doubt it as a *working hypothesis* in our considerations of Fermi's paradox.

But even this modest methodological naturalism can lead us into subtleties and difficulties concerning, for example, the role of philosophical inquiry in formulating synthetic hypotheses about the world. In the context of Fermi's paradox this might be interesting, since many of the proposed solutions are indeed located at the crossroads of science and philosophy and some of them could be regarded primarily as philosophical hypotheses. One such subtlety lies in assuming that the dichotomy between natural and supernatural applies to the 'internal' perspective of observers located within a given cosmological domain, or 'a universe'. The distinction is important for those discussions which do not question naturalism per se, but accept different construals of 'natural' depending on the 'internal' vs. 'external' perspective. For example, Max Tegmark argues in several papers for a radical ontology in which all possible mathematical structures are realized in some part of the (obviously huge) multiverse. Some parts, like ours, based on structures such as Hilbert spaces of quantum mechanics, support life and observers, which Tegmark calls 'self-aware substructures'. In this manner he emphasizes that there is a key difference between internal and external perspective, the former being what we as substructures are condemned to, the latter being a 'Platonic' view from outside of the realm of any particular physics.

What counts as naturalist explanation comes with (usually tacit) ontological commitments, especially where our best current physical theories are concerned

Obviously, from an internal perspective, agencies and capacities from a realm beyond one's own—possibly other evolved self-aware substructures, but now substructures of the *larger structure*—will not be limited by the local laws of physics, and thus, technically, could qualify as supernatural. Of course, this does not entail a real violation of the laws of physics at the highest level, that of Tegmark's 'ultimate ensemble theory'. However, it might *look* to any number of internal observers as if the laws of their local physics were violated. Although Tegmark's scenario is extreme, it does convey a general and important lesson that what counts as naturalist explanation—and looks like 'pure methodology' at first glance—comes with (usually tacit) ontological commitments, especially where our best current physical theories (such as string theory and inflationary cosmology) are concerned. This is in important lesson for considering some of the weirder candidates for the resolution of Fermi's paradox.

Of course, we could easily have a brand of supernaturalism that allowed Fermi's paradox to remain, and was entirely independent of it. Old pluralist views of the Creator endowing many stars with inhabited planets—or even, as Herschell believed, inhabited stars—and not having any particular preferences in the multitude of His creations, face perhaps even stronger forms of the paradox than do any naturalistic views discussed here. But that would violate an even more fundamental guideline: *usefulness*. Any solution to Fermi's question found in a completely naturalistic picture would still be valid if we allow for such a benign (relatively to the problem at hand) form of supernaturalism. Ultimately, supernaturalism is effectively a non-starter in discussions relating to Fermi's paradox. The very best one can do is simply to reject it. After all, its dubious virtue is that one can freely return to it if all else fails. At that stage, it might still have to compete, though, with the explanatory nihilism: the view that some deep questions simply have no answer and that phenomena to be explained are brute facts, neither requiring causal explanation, nor capable of being explained in any non-trivial way.

Scientific Realism

Realism has, as a term, been so widely used in the history of philosophy, denoting so many different things, that many openly refrain from its further (ab)use. Luckily enough, what we need here is not any deep meaning or general epistemological construal, but only the mundane observation that we achieve best results in most cases by following empirical evidence and interpreting it in the most direct and simple manner—not just in the scientific enterprise of explaining and predicting, but in other fields like sports, banking, or sex. Of course there are many examples of 'our eyes failing us' in both everyday life and in science and the arts, but their measure in the overall set of observations is quite small, at least until we reach very complex domains of theories such as quantum mechanics or general relativity (or works of artists at least partially inspired by those theories, such as M. C. Escher). I follow the Quinean recipe of using scientific knowledge as a yardstick for realism: the real is that which has been established by using empirical science, in particular methods such as observation and experiment. Thus, Uranus, inosilicates, *Pseudoceros dimidiatus*, or the Osaka airport are undoubtedly real, while unicorns, ghosts, black magic, or round squares are not. Many other things might be real, but we still do not know for sure; among such things are, of course, extraterrestrial intelligent beings. Note that this is an extremely weak understanding of realism, not referring at all to purely theoretical or abstract entities, making no commitment to the usual-suspect chestnuts such as mathematical objects or universals, making no claims about theories and their models. Rather, it is related to the aspirational attempts to define science: science aims to produce true descriptions of things in the world.

Realism in this weak sense is so central to the entire post-Renaissance scientific endeavour that there are very few cases in which it has been seriously questioned. And even in such cases, as in some more extreme subjective versions of the Copenhagen interpretation of quantum mechanics, it was not that the existence or *prima facie* validity of the empirical results obtained in the course of research was denied, but only our construal of the underlying physical reality. In contrast, non-realism in the context of Fermi's paradox requires that we reject the validity of a large amount of empirical data, amassed through decades or even centuries of careful work in observational astronomy.

Copernicanism

Copernicanism is usually the only philosophical element explicitly present in most discussions of

Copernicanism offers not only potentially testable hypotheses, but whole research programmes whose time is clearly yet to come

Fermi's paradox; that this is a dangerous oversimplification should not derail us from perceiving its central role in the paradox. Copernicanism is sometimes referred to as the 'Principle of Mediocrity' or something similar. This is misleading, and not only for the semantic reason that 'mediocre' is usually defined and understood as meaning 'of moderate or low quality'—which, obviously, has nothing to do with the astrobiological context. An even bigger problem is that it presupposes the ordering of elements in some well-defined way, which, even if possible in principle, is certainly not possible in any obvious way. 'The principle of typicality' would perhaps be a better designation, and indeed it has gained some currency in recent quantum cosmology.

If we accept Copernicanism, then, within reasonable temporal and physical constraints, we expect the status of biological evolution on Earth to reflect the Galactic average for given age of our habitat. We do not expect—unless we obtain specific reasons to the contrary!—that we have evolved exceptionally early or exceptionally late *in the interval within which the evolution of intelligence is physically possible*. In other words: the timescale for the evolution of intelligence on Earth is close to the median of the distribution of physically possible timescales for evolution of intelligence anywhere in the universe. This is clearly arguable, but for now we just note that Copernicanism offers not only potentially testable hypotheses, but whole research programmes whose time is clearly yet to come.

In the same manner, we have no reason to believe that the current astrobiological status of the Milky Way—whatever it may be—does not reflect accurately the average astrobiological status of the current universe, or at least its habitable subset. This is exactly the rationale for the assumption (widely used in the orthodox SETI since pioneering ideas of Iosif Shklovsky and Carl Sagan) that most of the members of the hypothetical 'Galactic Club' of communicating civilizations are significantly older than ours. This distribution reflects the underlying distribution of the ages of the terrestrial planets, coupled with a simple observation-selection effect: even slightly younger civilizations will not be detectable from interstellar distances, while older ones—if they do not go extinct—will be. Since our own civilization has essentially just appeared on the cosmic scene, it would make no sense whatsoever to search for a younger civilization (at present—things may become very different if humanity survives as a technological civilization for any astronomically relevant amount of time, like a few million years).

Exactly how much older than humanity do we expect the average Galactic civilization to be? Copernicanism alone will not be able to tell us that

Now, the intuitive impression may not accurately represent the quantitative reality. Exactly *how much* older than humanity do we expect the average Galactic civilization to be? Copernicanism alone will not be able to tell us that; neither will any other '-ism'. We need hard empirical data at this point. (Note that even just an order of magnitude would be a huge difference, and would make conventional SETI strategies rather problematic).

All of this fundamentally points to the importance of what Nikolay Kardashev calls 'initial concepts', in a paragraph that was written more than thirty years ago, but could be repeated almost verbatim today:

> Most experimental searches for extraterrestrial civilizations proceed from a position of 'Terrestrial Chauvinism'. Thus, in spite of criticism that the probability of finding a civilization at our level of development and—moreover—among the nearest stars is in fact close to zero, the search for Earth-type civilizations is continuing. The solution of the problem has not and will not be advanced until the initial concepts and therefore the search strategies are changed [...] Extraterrestrial civilizations have not yet been found, because in effect they have not yet been searched for.[2]

2. N.S. Kardashev, 'On the inevitability and possible forms of supercivilizations', in M. D. Papagiannis (ed.) *The Search for Extraterrestrial Life: Recent Developments* (Dordrecht: IAU, 1985), 497-504: 497.

What Kardashev calls 'initial concepts' are, of course, precisely philosophical concepts! This testifies, indirectly, to how poorly Copernicanism has been analysed in the astrobiological context so far—or even not taken seriously enough. The very fact that the first naive SETI 'models' of the founding-fathers epoch bluntly assumed that our civilization is close to typical testifies that, under a wide range of conditions, Copernicanism will be unthinkingly oversimplified. At the opposite extreme, contemporary 'rare Earth' theorists often unreflexively ignore or downplay Copernicanism, even when the level of our empirical understanding of the issues involved is obviously low and there is no reason whatsoever to doubt the Copernican assumption. So we need to navigate between the Scylla of naiveté and the Charybdis of simply ignoring the issue.

However, it is important to separate considerations of preference or even expediency from our analysis of what could we reasonably expect to be typical. Obviously, not all things we could, in principle, see, are typical. The fact that we observe no supercivilizations (of Kardashev's Type 3, for example) in the Milky Way in spite of ample time for their emergence is *prima facie* easiest to explain by postulating the vanishing probability or impossibility of their existence in general. On the other hand, if for some hitherto unknown reason the emergence of such supercivilizations lasts much longer, comparable to the Hubble time, than it is entirely Copernican to conclude that they are commonplace *averaged over all times*. Simply, some of the small civilizations existing now—humanity, for instance—could reasonably hope to become a supercivilization in future, and to stay in that state for a very long time.

An obvious pitfall for applications of Copernicanism in science in general, and astrobiology in particular, is to assume—explicitly or, as is almost always the case, tacitly—that the underlying distribution in parameter space is normal or Gaussian. This is usually wrong for some clear physical reason; for instance, the distribution of ages of terrestrial planets cannot be normal, since they are obviously bounded from above by the age of the thin disk. Just how big a mistake one makes in *assuming* a normal distribution around some mean value is, however, unclear in the general case. Unfortunately, in this unfamiliar context, the outliers are likely to be 'black swans':

rare occurrences with a tremendous and unforeseeable impact.

Of course, Copernicanism is just a principle: it cannot on its own ('in isolation') do the explanatory work for us. It needs to be coupled with correct empirical knowledge and theoretical ideas about the world. If it is coupled with incorrect ideas—or false empirical results—it will not help us, and may just confuse things more. A good example of such wrong-headed Copernicanism is the assumption, widely held in Victorian times and encapsulated in the internal logic of HG Wells's *War of the Worlds*, that, since Kant-Laplacean cosmogony implied that Mars is an older planet than Earth, a Copernican consequence is that Martian civilization must be technologically more advanced that the terrestrial one. The 'canals' of Mars were thus interpreted as global macro-engineering projects of this Martian civilization, giving proof for this Copernican hypothesis. Given the premise, the conclusion is quite reasonable: even more than a century later, we still know too little about the dynamics of cultural evolution to be able to put forward a cogent alternative model of technological development vs. time. But the premise (validity of the Kant-Laplace theory) was wrong, so the conclusions, which would nowadays be classified as belonging to SETI studies, were irrelevant. We need to be wary about encountering similar fallacies in the much more complex and sometimes counter-intuitive context of contemporary astrobiology.

Gradualism (and Red Herrings)

Gradualism, in the sense that we shall use it here, is brilliantly captured by Lyell's slogan 'the present is key to the past': gradualism essentially suggests that, in explaining the features of the past record, we should take into account only processes observable by our systematic inquiry (at present, since all observation happens at epochs close enough to be for all practical purposes simultaneous).[3] In other words, we need not postulate any epochs substantially different from the present one in terms of acting forces and processes. If we seem to observe such an epoch, this must be owing either to our incomplete understanding of present processes—which is always on the cards—or to our lack of theoretical sophistication in building an explanation based only on present processes.

How has it come to pass that gradualism is treated as a philosophical assumption, and not as a straightforward question to be empirically resolved one way or another?

In fact, the present-day position is such that one could ask an almost inverse question: after so many empirical developments in Earth sciences and around them, after so many debates in learned scientific journals, starting with Hutton's *Theory of the Earth*, how has it come to pass that gradualism is treated here as a *philosophical* assumption (or principle), and not as a straightforward question to be empirically resolved one way or another?

Actually, the situation is somewhat more complicated. Gradualism is regarded as a philosophical principle here, rather than a straightforward empirical matter, since there have been many historical ambiguities surrounding its definition, resolution and the domain of applicability. It is obvious that it means different things for different people and different things in different epochs. That part of the analysis clearly necessary for disentangling various nuances of meaning and doctrine must be epistemological and methodological goes without saying. That a particular instance of the general phenomenon X occurred in small increments over long period of time is an empirical claim. If empirical research tells us that this is indeed the case, we might conclude that *this particular instance of X was gradual*, and it might give us some confirming evidence about the gradual nature of X *in general* although just how weighty such a piece of evidence is depends on further assumptions about probability, inference, and justification). To prove the general claim for X in all instances is a completely different and perhaps unfeasible task; and in the astrobiological context, the problem is compounded by the fact that one reasonably expects the diversity of possible contexts in which instances of X arise to be much larger than what has traditionally been considered in the geosciences.

For example, the frequency and strength distribution of earthquakes on Earth are, at least in principle, determined by details of plate tectonics, which in

3. Obviously, this does not apply to cosmology.

Timescales are invented by humans and are cultural artefacts. Planetary and Galactic history do not impose any particular preferred timescales

turn are set by global parameters fixed at the time of accretion of Earth, 4.556 Ga ago; such global parameters are the total mass and density of the planet, and its chemical composition, especially the amount of long-lived r-nuclides such as ^{238}U and ^{232}Th. On a different terrestrial planet in the Milky Way where those parameters are different, the frequency and the range of intensities of quakes will be different; and the hypothetical inhabitants of such a planet might draw vastly different conclusions about the gradual nature of quakes which, on Earth, although often *catastrophic* in their consequences, have been both local in terms of consequences, and happen often enough to be subsumed under the Lyellian 'present' (which is the key to the past). Therefore we on Earth do not consider them a vindication of catastrophism. But on an Earth-like planet where quakes happened only once per century or even more rarely, but always with high intensity, would they be considered a gradual phenomenon? Therefore, an Achilles' heel of gradualism is its dependence on non-uniquely defined timescales; such timescales are invented by humans and are cultural artefacts, rather than something built into nature itself. Planetary and Galactic history do not impose any particular preferred timescales. Thus, whether we estimate the probability of a decisive catastrophic event per year or per Ma or per Ga should not influence our conclusions and their interpretation—but, obviously, gradualism on longer timescales looks pretty much 'punctuated', and its explanatory power is greatly diminished.

As distinguished paleontologist David M. Raup writes:

> An interesting aspect of the history of meteorite study is that the uniformitarian doctrines of Lyell and his followers have been able to absorb the new facts and concepts without seriously changing the basic catechism. Meteorites fall and they make craters, often big ones. Once accepted, this pattern became part of the uniformitarian doctrine and was no longer considered catastrophic.... One can argue, therefore, that catastrophism really means something that is unfamiliar. As soon as it becomes familiar, the awful label need no longer burden it. Scientists are adaptable people.[4]

And if this is so complex an issue when we are dealing with geological phenomena, which have been the playground for the debates surrounding gradualism since Hutton's time, how much more complex must it be when applied to the emergence and evolution of intelligence and technology throughout the universe! That very complexity justifies our temporary treatment of gradualism as a philosophical principle, in substantially the same manner as the completeness of formal systems was treated before Gödel's great discovery in 1931. (This does not presume that gradualism will be eventually proven wrong, just that it is too complex an issue for our present modest understanding of the problem-situation.)

The Non-Exclusivity Principle

In order to tackle what has been a source of almost infinite confusion in countless debates on Fermi's paradox and on SETI in general, let's turn to sports. If a runner such as Husain Bolt runs well, it is mostly due to his actual prowess, allowing for some rather minor impact of weather conditions or other extraneous factors, as well as of what might be called the historical record: his trainers, previous work, etc. But when a team like Arsenal plays badly and loses, on the other hand, the spectrum of possible reasons for this is very broad indeed, for it is very difficult in practice to really determine the causal structure of any team event. But suppose that we have a kind of 'Laplace demon' insight into the causes of each particular match outcome in terms of any individual player's performance. For example, in the Arsenal-Aston Villa 5–0 result (1 February 2015, The Emirates Stadium), our insight tells us that the performance of the goalkeeper Ospina was 75%, the striker Giroud 90%, etc. Would this be enough to give the causal account of the outcome of the entire championship (say the Premiership or the Euroleague)? And, in particular, to what extent would the same causes underlying the outcome of each individual

4. D.M. Raup, *The Nemesis Affair: A Story of the Death of Dinosaurs and the Ways of Science* (2nd edition, New York: W. W. Norton, 1999), 35–36.

game determine the final outcome of a series of matches comprising such 'higher-order' event? If a team A wins the championship in spite of mediocre performance in most matches, shall we claim that there are *additional* causes of this unexpected success, over and above the usual causal properties? Can occasional 'flashes of brilliance' confer the cup on an average team in a prolonged championship? Such questions become particularly interesting and even dramatic if we introduce a situation in which, for some objective reason, say family or job troubles, we cannot study *all* matches in the season, studying instead—with the same perfect insight in each individual case—just a more or less representative sample of the season's matches.

The sporting analogy is apt for an additional reason, the importance of which will soon become clear: elements of regularity (baseline quality of players and the coach, enthusiasm of fans or the lack of it, etc.) usually persist throughout a season, on timescales much longer than that of any single match. On the other hand, a season consists of individual matches—any positive or negative fluctuation affects the average. So is constancy really more important for winning cups than what we might call a potential for surprises? In spite of the pundits, it is very hard to tell in a complex environment such as football. And the same goes for astrobiology, where it is even harder to pinpoint where exactly the surprises may come from. But it is reasonable to conclude that, lacking deeper insight, constancy of form is preeminent.

Hence, we need another general principle. Non-exclusivity was introduced and elaborated upon by David Brin in his seminal 1983 study of Fermi's paradox. He writes:

> [W]e are tempted to add one more, rather tentative, 'principle' [...] a 'Principle of Non-exclusiveness', which states that diversity will tend to prevail unless there exists a mechanism to enforce conformity.[5]

Subsequently, Brin demonstrates how this works on many examples of particular hypotheses posed to explain Fermi's paradox:

> Non-exclusiveness would seem to apply to Resource Exhaustion. Even if some ETIS were wastrels, at least a few others would see the crunch coming and plan for it [...] To overcome non-exclusivity we need a mechanism which might affect fc systematically, so that there are few if any exceptions to slip away and fill the Galaxy with the commerce we do not observe.[6]

If there is a single causal mechanism acting on all civilizations over all of the history of the Galaxy, we would have a properly non-exclusive explanation of the observed Great Silence

As another oft-encountered example, consider the classical hypothesis (let us call it A) that all cosmic civilizations self-destruct upon developing nuclear or biological weapons. It is quite exclusive: it assumes that the same global outcome arises independently from a wide variety of local properties. The same disastrous history, the same global holocaust gets repeated over and over again, on millions of planets separated by kiloparsecs and millions of years. This is intuitively improbable, and Brin's principle of non-exclusivity gives a formal statement of such intuitions. If, however, there is a single causal mechanism acting on *all* civilizations over *all* of the history of the Galaxy, we would have a properly non-exclusive explanation of the observed Great Silence. For instance (hypothesis B), if the very first Galactic civilization actually quarantines all the subsequent ones, including humanity, into local 'zoos', taking care to isolate them from the others, then we have a single causal explanation for non-observation of many different Galactic civilizations and their activities, as well as the fact that none has visited Earth and the Solar System thus far. It does not matter, from our present standpoint, that we might find either hypothesis A more probable or hypothesis B more improbable *on other grounds*—it is just the illustration of an *additional desideratum* which the principle of non-exclusivity offers us. (Indeed, B was conceived specifically for this purpose out of several related far-fetched explanatory hypotheses for

5. G. D. Brin, 'The "Great Silence": the Controversy Concerning Extraterrestrial Intelligence', *Q. Jl. R. astr. Soc.* 24 (1983), 283–309: 287.

6. Ibid., 296. 'ETIS' stands for 'extraterrestrial intelligent species'. fc is a factor in the Drake equation (see below) representing the fraction of the intelligent species which develop detectable technologies.

Fermi's paradox, while A has a long and venerable tradition in SETI studies.) As with other philosophical criteria considered here, it makes little sense to use non-exclusivity alone for judging the solutions to the puzzle; we need all—or most—of them in order to arrive at the really strong contenders.

Now, isn't non-exclusivity just a new-fangled name for good old Occam's razor? In fact this is not the case, for the following two reasons. Occam's razor, as it is usually construed, talks about the simplicity of explanatory hypotheses (or assumptions built in such hypotheses) and not about the scope of such hypotheses. In the usual and mundane circumstances of our terrestrial lives, it is easy to conflate the two, since complexity and scope are often correlated: it is certainly easier to understand and explain the economic functions of a bakery on your corner than those of the London Stock Exchange. However, there is no such correlation on the astrobiological scene. If a simple probabilistic cause acts uniformly on 109 terrestrial planets in the Galaxy in the course of various epochs (as in hypothesis A above), we have no rational reason to prefer it over a complex cause which acts once for all time and achieves the same explanatory purpose (as given by hypothesis B).

The other reason to resist unthinking identification of the non-exclusivity principle with Occam's razor is that Occam's razor properly applies only to situations in which the relevant assumption is *ceteris paribus*: all other things being equal. Occam's razor suggests that, after we all agree on the empirical facts about how combustion proceeds in the real world, it is better to explain the phenomenon of combustion by oxidation than by release of the hitherto unknown mysterious substance called phlogiston. In other words, the hypothesis of oxidation is better than the hypothesis of phlogiston since it is simpler *ceteris paribus*. No *ceteris paribus*, no Occam's razor. And this is likely to cause tremendous difficulties for application of Occam's razor in studying Fermi's paradox, since it involves comparisons of the evolutionary status of hypothetical wildly varying biospheres from all over the Galaxy and over billions of years of cosmic time. The prospects of having a common *ceteris paribus* in such a situation are non-existent. Moreover, it is doubtful that ceteris paribus is applicable to hypotheses for explaining particular features of biological (macro)evolution even on a single planet, Earth!

For these reasons, I maintain that we need to consider non-exclusivity as a specific criterion for the evaluation of hypotheses for explaining Fermi's paradox, independently of Occam's razor.

Non-exclusivity is a particular application of a wider principle of logic, which rejects solutions that are worse than the problem itself. Australian astronomer Luke Barnes calls such solution cane toad solutions, from an ecological example too beautiful not to be quoted:

> In 1935, the Bureau of Sugar Experiment Stations was worried by the effect of the native cane beetle on Australian sugar cane crops. They introduced 102 cane toads, imported from Hawaii, into parts of Northern Queensland in the hope that they would eat the beetles. And thus the problem was solved forever, except for the 200 million cane toads that now call eastern Australia home, eating smaller native animals, and secreting a poison that kills any larger animal that preys on them. A *cane toad solution*, then, is one that doesn't consider whether the end result is *worse* than the problem itself.[7]

In the context we are discussing here, such cane toad solutions would all require an improbable conspiracy of causes, acting over long intervals of time and a huge volume of cosmic space, to produce the observed Great Silence. The *Hermit Hypothesis*—that intelligent beings never expand beyond their home planetary system, communicate, or in any other way become detectable over interstellar distances—is one such cane toad solution, which violates the non-exclusivity principle with gusto. Therefore it is not really a serious candidate solution to Fermi's paradox: at best, it requires an improbable conjunction of causes. As usual, one can return to it—and other hypotheses violating non-exclusivity—*if all else fails*; but we had better not hurry there.

The Hermit Hypothesis is not really a serious candidate solution to Fermi's paradox: at best, it requires an improbable conjunction of causes

7. L.A. Barnes, 'The Fine-Tuning of the Universe for Intelligent Life', *Publications of the Astronomical Society of Australia* 29 (2012), 529–564: 531.

Likewise, we can regard those solutions requiring that we give up solid scientific evidence or even the scientific method itself as particularly egregious cane toad solutions.

One thing is crucial here: non-exclusivity might not be *true* (in the same sense that vast majority of scientists hold that naturalism or realism are true, being reliable guides to the scientific truth in any particular matter of investigation), but we still have solid reasons for using it as a tool for sorting out the jungle of hypotheses and performing the taxonomical task at hand. In conjunction with the other philosophical criteria outlined above—realism, naturalism, copernicanism, and gradualism—it forms a set of desiderata for the true solution of the big puzzle.

However, since Fermi's paradox is a specific scientific problem, one which will ultimately be solvable by empirical means (if humanity does not destroy itself first), and it is a *contingent fact of the (astrobiological) history of the Milky Way*. Therefore, non-exclusivity is not on a par with metaphysical assumptions (such as realism) or epistemological assumptions (such as scientific naturalism); nor it is dependent (as gradualism is) on the spatial and temporal framework we choose to work in. As is the case with Copernicanism, we can reasonably hope to establish the truth value of non-exclusivity by future research, and to explain why it has one truth value or the other. However, since that prospect is not immediate, and perhaps lies in quite distant future, I feel justified in treating it here as a philosophical assumption.[8]

The Continuity Thesis

Iris Fry showed that a necessary ingredient in any account of abiogenesis (and, by extension, the origin of intelligence, or *noogenesis*) in naturalistic terms is the so-called continuity thesis: 'the assumption that there is no unbridgeable gap between inorganic matter and living systems, and that under suitable physical conditions the emergence of life is highly probable.' Adherence to the continuity thesis, as Fry amply demonstrates, is a precondition for scientific study of the origin of life; contrariwise, the views that abiogenesis is a 'happy accident' or 'almost a miracle' are essentially creationist, i.e., unscientific.

8. There are rough analogies to this situation in the history of science. For instance, functionalism with respect to mechanisms of biological evolution was rather an abstract philosophical assumption in the times of Lamarck and Darwin—who both were ardent functionalists. It was an excessively fruitful philosophical assumption, which by far most biologists in subsequent both darwinist and lamarckist traditions believed to be empirically true, but it could not be proved as such (in its proper domain, of course) until the advent of molecular biology.

While the naturalist worldview implies that life and observers like ourselves are compatible with the laws of physics, there is nothing in the laws themselves which tells us the frequency of such events as abiogenesis or noogenesis

The continuity thesis has nothing to do with a metaphysical notion of *necessity*. It does not state that the emergence of life or intelligence is necessary in any reasonable construal of the word. Nor does it claim that we can derive features of emerging life or intelligence or technological civilizations from the form of the laws of physics (even if they were known fully). While the naturalist worldview implies that life and observers like ourselves are compatible with the laws of physics, there is nothing in the laws themselves which tells us the frequency of such events as abiogenesis or noogenesis. The continuity thesis is a working hypothesis (as well as a heuristics, as explained above) that the part of the relevant parameter space containing life or intelligence is appreciably large; it might be wrong, but it certainly is not tautologous or vacuous—or non-empirical. In short, it is philosophical!

Postbiological Evolution

The necessity of taking into account the possibility of *postbiological* evolution has only been recognized rather recently, prompted by Moore's Law and the great strides made in the cognitive sciences. There is much talk now—if not always in scientific, but at least in proto-scientific terms—about different forms of human enhancement, implants, cyborgs, mind uploading, or even technological singularity. While many people continue to consider these topics as staples of science-fiction, they are fighting the rear-guard now: the future is coming and,

We should not expect other features we observe in human, still largely biologically grounded, society to be truly universal, that is to apply to postbiological/nonbiological societies as well

barring some global cataclysm and destruction of human civilization, these and related topics will certainly become more active and important as science and technology progress.

It is easy to understand the necessity of redefining SETI studies in general and our view of Fermi's Paradox in particular in this context. For example, postbiological evolution makes those behavioural and social traits such as territoriality or expansion drive (to fill the available ecological niche) which are—more or less successfully—'derived from nature' lose their relevance. Other important guidelines must be derived which will encompass the vast realm of possibilities stemming from the concept of postbiological evolution.

Since the 'expansion drive' has often been promoted as the root of all problems with Fermi's paradox, and given that naive readings of it ('you certainly do not expect advanced extraterrestrial intelligences to behave like the Spanish conquistadors?', with all the appropriate body language and gestures of offence) still occupy a large niche in the literature, it is worth going here into slightly more detail. Without entering the debate on merits or demerits of sociobiological and evolutionary-psychological explanations, one thing is clear on purely logical grounds: if we accept that *in principle* these disciplines could be satisfactorily grounded in some inclusive form of biological evolution (a very weak assumption), and we accept that postbiological evolution is a serious possibility (another weak assumption), then it is clear that we should not use concepts derived from our biological past to explain features of our postbiological future. Ditto for other hypothetical intelligent species in the universe. In the specific case, very relevant for Fermi's paradox: we should not expect that postbiological civilizations will have the same evolutionary-conditioned 'expansion drive' as the biological ones. We should not expect other features we observe in human, still largely biologically grounded, society to be truly universal, that is to apply to postbiological/nonbiological societies as well; particularly important among those features are aggression and warfare, both of which are claimed by sociobiologists, with varying success, to follow from our evolutionary inheritance. Neither should we consider the future of humanity or the advanced stages of evolution of other intelligent beings to be bounded by the Malthusian problems, which, as is well-known, motivated Darwin himself. In brief: we should avoid biases grounded in our biological origin.

There seems at present to be no conceivable reason why postbiological evolution should not be possible. Whether it will come to pass on Earth—or in any other place inhabited at any moment of time by intelligent beings—is a completely different question. To claim that humanity is either a very good or a very bad model for possibilities of postbiological evolution is to violate Copernicanism without any compelling reason to do so. However, if some civilizations go along the postbiological pathway and some not, it is reasonable to expect an outcome which can be construed as a monumental irony: the mechanism of *natural selection*, that keystone of all *biological* approaches to explanation, might give a huge advantage to *nonbiological* actors on the Galactic stage.

This may be said to follow from a sobering fact mundane to anyone acquainted with technology (and history) or astronautics: Machines are, simply, much better at tasks relevant for the exploration and colonization of space than people *in their current biological form*.[9] It is exactly for this reason that Tipler's scenario with self-replicating interstellar probes is the strongest form of Fermi's paradox and the biggest obstacle any contact-optimist has to face.

In short, it is important to keep the postbiological option open as we consider each and every of the explanatory hypotheses for Fermi's paradox.

The Drake Equation, For Good or Bad

It seems impossible even tangentially discuss SETI issues without the customary, or even ritual, mention of the Drake equation, which supposedly

9. Note that *this* apparent violation of Copernicanism is indeed (trivially) justified by our empirical knowledge on the prevailing conditions in the Solar System, say: temperature and pressure extremes, lack of atmospheric oxygen, etc.

supplies us in principle with an valid measure of the number of technological civilizations likely to exist in the universe. Somewhat paradoxically, this does not have to do with the substance of the discussion; in some cases, it seems to be motivated exclusively by the fear that a discourse without equations will not be taken seriously enough in the 'real' scientific world. Such idolatry of mathematics and numbers is entirely misplaced; there have been many similar grotesque cases of trying to force mathematical language and formalism on fields such as literary criticism or art history or class struggle. Even in the areas where mathematical expression gradually took roots, such as evolutionary biology, this occurred due to deep methodological and historical reasons, and not because practitioners felt insecure and uncomfortable without repeating *ad nauseam* some simple piece of mathematical regularity. In the SETI field, invocation of the Drake equation is nowadays largely an admission of failure. Not the failure to detect extraterrestrial signals—since it would be foolish to presuppose that the timescale for the search has any particular a priori range of values, especially with such meagre detection capacities—but the failure to develop the real theoretical grounding for the search. This follows from its very structure:[10]

$$N = R^* \cdot f_p \cdot n_e \cdot f_l \cdot f_i \cdot f_c \cdot L$$

where N is the 'predicted' number of extraterrestrial civilizations (usually misinterpreted as the number of SETI targets, see below) in the Milky Way; R^* is the star-formation rate in the Milky Way, appropriately averaged; f_p is the fraction of stars possessing planets of any kind; f_l is the average number of habitable planets per planetary system; n_e is the fraction of habitable planets actually possessing life (either through abiogenesis or panspermia); f_l is the fraction of inhabited planets developing intelligent life; f_i is the fraction of intelligent communities developing the technology relevant for detection and communication over the interstellar distance; and the (in)famous factor L is the lifespan of the civilization in the detectable mode (again, often misinterpreted in a variety of ways).

Each fraction (or probability) term in the Drake equation should, in fact, be explicated in terms of relevant probability distribution functions, integrated over the relevant region in the parameter space. Roughly speaking, there should be an integro-differential equation for each of the probability terms in the equation. For instance, the fraction of habitable planets is an integral over the rate of planets becoming habitable minus the rate of their ceasing to be habitable by various processes (runaway greenhouse effect, stellar evolution, etc.); this should be integrated over spatial volume of the Galaxy over the course of the Galactic history. In this manner, we could finally make the transition between guessing (even educated guessing!) and computing within the context of a quantitative astrobiological models. In fact, all individual terms, with a partial exception of the average lifetime of technological societies, belong squarely and without any doubt into the domain of astrobiological research, which becomes more and more sophisticated and precise as we speak, write or read these lines. So, faced with those who continue to rewrite, cite, and re-cite the Drake equation without trying to get a deeper theoretical insight, one is more and more justified to say 'put up—or shut up!'

Faced with those who continue to rewrite, cite, and re-cite the Drake equation without trying to get a deeper theoretical insight, one is more and more justified to say 'put up—or shut up!'

The tendency to portray the Drake equation as a cornerstone—or, even worse, the cornerstone—of SETI by both proponents and opponents alike should give one pause. After all, in other grand controversies in the history of science, a particular piece of theoretical apparatus was invoked by either critics (as a weakness) or defenders (as a strength), but not by both alike. Consider, for instance, epicycles in the old geocentric cosmology of Ptolemy: they were introduced to 'save the phenomena' and no rhetorical manoeuvre could make them a *virtue* of the theory. The Copernican opponents charged that epicycles are one of the major *weaknesses* of the old theory; supporters, such as Clavius or Riccioli, at best muttered something along the line of 'they enable us to make accurate predictions'. It would be a strange situation indeed if geocentrists were to emphasize

10. F. Drake, 'The Drake Equation Revisited: Part I', *Astrobiology Magazine*, 29 September 2003.

epicycles as one of their great achievements! Such examples abound in history of science, and all contrast sharply to what we have seen in more than half century of contemporary SETI debates.

Therefore, it is high time for the practice of ritual invocation of the Drake equation to stop. While the SETI research community cannot, of course, prevent abuses of history and its 'Whiggish' reinterpretation, what it can do is to start insisting on building more a serious theoretical scaffolding for its enterprise. Since the Drake equation is a rule of thumb which should be derivable from any such real SETI theory by a chain of approximations, integrations, and averaging, it should not be advertised as anything more than that. Instead, we should strive to reach deeper understanding through more precision, more numerical models, more simulations, more specific scenaria subject to quantification, etc. Fortunately, the winds seem to be changing recently in this area as well.

*

As the history of science teaches us, philosophical considerations (often disguised as aesthetical or ethical judgments, bold metaphors, or 'leaps of imagination') have always played a key role in the early stages of any scientific discipline. This was the case with chemistry in the time of Lavoisier, geo-sciences in the time of Hutton, or cosmology in the time of Friedmann and Lemaître. Subsequently, those things might fade in the background—as they did in cosmology, for instance, after the 'great controversy' between the standard relativistic cosmology and the steady-state alternative was over in the mid-1960s—and become remote from everyday work of practicing researchers, but they did play a crucial role, as even a cursory non-myopic insight into the basic tenets and concepts of the field will reveal. Likewise, in these early years of astrobiology and SETI, we must recognise the importance of the philosophical principles outlined above in guiding our inquiry and developing our insights.

Milan Cirkovic's book *The Great Silence: Science and Philosophy of Fermi's Paradox* will be published by Oxford University Press in 2017.