Serb. Astron. J. № 166 (2003), 1 - 11

UDC 52–37
Original scientific paper

EXTRATERRESTRIAL INTELLIGENCE AND DOOMSDAY: A CRITICAL ASSESSMENT OF THE NO-OUTSIDER REQUIREMENT

M. M. Ćirković[1] and V. Milošević-Zdjelar[2]

[1] *Astronomical Observatory, Volgina 7, 11160 Belgrade 74, Serbia and Montenegro*

[2] *Physics Department, University of Winnipeg, Winnipeg MB, R3B 2E9, Canada*

(Received: January 23, 2003; Accepted: March 14, 2003)

SUMMARY: The so-called Doomsday Argument has attracted a lot of attention in recent years. One of the suggested ways to make this argument ineffective is the so-called No-Outsider requirement, indicating that the presence of other intelligent observers ('aliens') beside humans invalidates the reasoning leading to the apocalyptic conclusion. Obviously, this argumentation bears relevance not only to the effectiveness of the Doomsday Argument, but also to the issues like the relativization of the reference class in anthropic reasoning, transhumanism and the SETI theory. Hereby we criticically investigate the No-Outsider Requirement and conclude that it is either fallacious or irrelevant.

Key words. extraterrestrial inteligence – history and philosophy of astronomy

1. INTRODUCTION: ANTHROPIC REASONING AND THE DOOMSDAY ARGUMENT

The Doomsday argument (henceforth DA) was first formulated some fifteen years ago when Brandon Carter found a hitherto unnoticed consequence of the weak anthropic principle. Carter did not publish his finding, but the idea was taken up and further developed by John Leslie (e.g. 1992, 1993, 1996) and Richard Gott (1993, 1994), who had discovered the argument independently. Roughly, the Doomsday argument reasons from our temporal position according to a principle directly analogous to the one followed by the other applications of anthropic reasoning from the expected typicality of our position in the multiverse or our spatial position within a universe.

The core idea can be expressed through the following urn ball experiment. Assume that two large urns are put in front of you, and you know that one of them contains ten balls and the other a million, but you are ignorant as to which is which. The balls in each urn are numbered 1, 2, 3, 4 ... etc. Now you take a ball at random from the left urn, and it shows the number 7. Clearly, this is a strong indication that that urn contains only ten balls. If originally the odds were fifty-fifty (identically-looking urns), an application of Bayes' theorem gives you the posterior probability that the left urn is the one with only ten balls as P_{post} (n=10) = 0.99999. But now consider the case where instead of the urns you have two possible models of humanity, and instead of balls you have human individuals, ranked according to birth order. One model suggests that human race will soon go extinct (or at least that the number of individuals will be greatly reduced), and as a consequence, the total number of humans that will have ever existed is about 10^{11}. The other model indicates that humanity will colonize other planets, spread through the Galaxy, and continue its existence for many future millennia; consequently, we can take the number of humans in this model to be of the order of, say, 10^{18}. As a matter of fact, you happen to find that your rank is about sixty billion. According to Carter and Leslie, we should reason in the same way as we did with the urn balls. To have a rank of sixty billion or so is much more likely if only 100 billion persons will ever have lived than if there will be 10^{18} persons. Therefore, by Bayes' theorem, you should update your beliefs about mankind's prospects and realize that an impending doomsday is much more probable than you have previously thought.

Leslie, in particular, has written on the topic at great length and has clarified the Bayesian structure of the argument and defended it against a range of philosophical objections and misconceptions. In recent years there has been a surge of interest in the DA; e.g. Bostrom (1999, 2000, 2001, 2002a); Kopf, Krtous and Page (1994); Olum (2002); Bostrom and Ćirković (2003). However, most of the publications have been in journals of philosophy, and the discussion has so far not connected closely to parallel debates among cosmologists and astrobiologists.

This state of affairs is unsatisfactory, not only because the DA is concerned with the same subject matter (the place and future prospects of intelligent life in the universe) but also because the Doomsday argument is an application of the very same idea of observer randomness that is an integral part of anthropic reasoning in cosmology (Leslie 1989b; Bostrom 2002a). The Doomsday argument is therefore not an optional additional consideration that can be brought to bear on the issue, but a direct consequence of an assumption already made in discussions about how the world-ensemble theories (such as many versions of inflationary cosmology) might explain the appearance of fine-tuning in the observable cosmos. It is a **consistency requirement**, therefore, that the Doomsday argument is to be taken into account in any application of anthropic reasoning. In particular, it is so, when the subjects of our investigation are actual prospects of survival of humanity in the framework of realistic cosmological models. The latter topic has generated an enormous interest among astronomers and philosophers alike in last several decades, as a part of the nascent discipline of physical eschatology (e.g. Adams and Laughlin 1997, 1999; Tipler 1986; Ćirković and Bostrom 2000; Oppy 2002).

2. A CHEAP SOLUTION: NO-OUTSIDER REQUIREMENT

A rather simple solution to the DA problem has been proposed by Bostrom (2002a), under the name of *no-outsider requirement* (henceforth NOR):

> The second reason for the doomsayer not to grant a probability shift in the above example is that the *no-outsider requirement* is not satisfied. The no-outsider requirement states that in applying SSA there must be no outsiders – beings that are ignored in the reasoning that really belong in the reference class. Applying SSA in the presence of outsiders will yield erroneous conclusions in many cases.

Bostrom extensively uses NOR in dealing with the objections to DA, but also in an attempt to show that DA reasoning is formally valid, but should not worry us seriously, since in the real world none of the assumptions behind it is satisfied.

What justification does Bostrom offer for NOR? Although not explicitly obvious, we can find some arguments in the following analyses:

> Consider first the original application of DA (to the survival of the human species). Suppose you were certain that there is extraterrestrial intelligent life, and that you know that there are a million "small" civilizations that will have contained 200 billion persons each and a million "large" civilizations that will have contained 200 trillion persons each. Suppose you know that the human species is one of these civilizations but you don't know whether it is small or large.
>
> To calculate the probability that doom will strike soon (i.e. that the human species is "Small") we can proceed in three steps:
>
> *Step 1.* Estimate the empirical prior Pr(*Small*), i.e. how likely it seems that germ warfare etc. will put an end to our species before it gets large. At this stage you don't take into account any form of the Doomsday argument or anthropic reasoning.
>
> *Step 2.* Now take account of the fact that most people find themselves in large civilizations. Let H be the proposition "I am a human." And define the new probability function $\Pr^*(\,.\,) = \Pr(\,.\,|H)$ obtained by conditioning on H. By Bayes' theorem,
>
> $$\Pr^*(Small) = \Pr(Small|H) = = \frac{\Pr(H|Small) \times \Pr(Small)}{\Pr(H)}.$$
>
> A similar expression holds for $\neg$ *Small*. Assuming you can regard yourself a random sample from the set of all persons, we have
>
> $$\Pr(H|Small) = = \frac{200\,billion}{(200\,billion + 200\,trillion) \times 1\,million}, \text{ and}$$
>
> $$\Pr(H|\neg\,Small) = = \frac{200\,trillion}{(200\,billion + 200\,trillion) \times 1\,million}.$$
>
> (If we calculate $\Pr^*(Small)$ we find that it is very small for any realistic prior. In other words, at this stage in the calculation, it looks as if the human species is very likely long-lasting.)

Step 3. Finally we take account of DA. Let E be the proposition that you find yourself "early", i.e. that you are among the first 200 billion persons in your species. Conditioning on this evidence, we get the posterior probability function $\Pr^{**}(\,.\,) = \Pr^{*}(\,.\mid E)$. So

$$\Pr^{**}(Small) = \Pr^{*}(Small|E) = \\ = \frac{\Pr^{*}(E|Small) \times \Pr^{*}(Small)}{\Pr^{*}(E)}.$$

Note that $\Pr^{*}(E \mid Small) = 1$ and $\Pr^{*}(E \mid \neg Small) = 1/1000$. By substituting back into the above expressions it is then easy to verify that

$$\frac{\Pr^{**}(Small)}{\Pr^{**}(\neg\, Small)} = \frac{\Pr(Small)}{\Pr(\neg\, Small)}.$$

We thus see that we get back the empirical probabilities we started from. The Doomsday argument (in Step 3) only served to cancel the effect which we took into account in Step 2, namely that you were more likely to turn out to be in the human species given that the human species is one of the large rather than one of the small civilizations. This shows that if we assume we know that there are both "large" and "small" extraterrestrial civilizations – the precise numbers in the above example don't matter – then the right probabilities are the ones given by the naïve empirical prior.

Status of NOR is highly uncertain, since it looks more than almost anything else in science and philosophy like a *non sequitur*. At first glance, the existence of aliens should not have anything to do with the fate of humanity. However, the issue here is not the fate of humanity as such, but rather the formal validity of reasoning leading to DA. Thus, NOR has some chances of success. It is our goal in this paper to show that these chances are rather slim, and that here the first glance is the correct one.

We shall try to reach this goal along two main lines. First, we shall attempt to investigate the other possible sources of NOR and to demonstrate their *non sequitur* or fallacious nature. We shall demonstrate that NOR is truly closely related to several conjectures and arguments appearing in the literature, like Carter's anthropic argument or rather controversial Self- Indication Assumption (henceforth SIA), which are sources of much confusion in anthropic reasoning. In addition, there seems to be some sort of wishful thinking, and partly sociological pressure leading to some of the reasoning behind NOR.

Second, we wish to point out that NOR fails to take into account very important cosmological facts, notably the existence of **cosmological horizons**, as well as new data on the size and geometry of the universe. When these are understood properly, they give us superb reasons for rejection of NOR, at least in its unrestricted form given above. Jumping somewhat ahead of the discussion, we may wish to state our opinion that there are only two real issues here: (i) whether there are ETIs within our cosmological horizon, and (ii) whether it is the nature of our domain such that the reference class may be defined at the superhorizon scale or not. We may wish to consider two cases here: the universe with only visual horizon, and that with both visual and event horizons.[1] At least in the case of a constant event horizon we intuitively seek to restrict NOR to those "outsiders" located within our horizon. However, Bostrom (private communication) does not accept such restriction, and his assessment is repeated, albeit in critical terms, by Olum (2002). Unrestricted in this way, NOR does not tell us anything useful or interesting, since we may be certain that $N_{total} = \infty$, but the problem is that the restriction is not unique. We shall return to this point in the discourse below.

Finally, we comment upon the relationship of DA and NOR with one of the most interesting and fascinating enterprises of modern age, which is SETI (Search for ExtraTerrestrial Intelligence). We suggest that, in the spirit of Gott's original argument, DA gives us a consistency argument for explanation of the so-called Fermi's paradox, which plays a crucial role in theoretical SETI studies. This, of course, is antithetical to the spirit of NOR, and presents a further reason for its rejection. To paraphrase Einstein's well-known judgement on Bohm-de Broglie quantum theory in a letter to Born (e.g. Bell 1987, p. 91), this solution to a DA problem "seems too cheap".

3. MARS ATTACKS?

Let us consider a counterfactual world–rather close to the actual one in terms of Lewis' theory of counterfactuals (e.g. Lewis 1986)–in which Mars is inhabited by a sophisticated technological civilization in the manner of H. G. Wells. As is well-known, Martians in Wells' story are slowly declining in numbers due to the long-term climatic change on their native planet. Suppose that their current population is rather small, say 10^6 or so, much smaller than the current human population. Would we have taken NOR seriously in such a world? Could we argue that DA reasoning is invalid because a couple of million of Martians live several dozens million miles away from Earth? Hardly. Instead, while pondering issue of past histories and relationship to future, we could choose one of the two major options. We could (i) consider the existence of Martians as *non sequitur* (which is certainly a usual, intuitive reaction), the irelevant but admissible piece of information, something similar to the fact that we observe the gravi-

[1] Of course, there may be a combination of the two, like in the models with decaying cosmological constant.

tational microlensing optical depth towards Large Magellanic Cloud to be several times 10^{-7}. Otherwise, we could (ii) simply add terrestrial and Martian populations at all times, and then draw our Bayesian (i.e. "doomslike") conclusions. In the particular case, small Martian population would not essentially change the conclusions of the doomsyear, as can be easily checked by elementary calculation.[2] In both options (i) and (ii) NOR is refuted, although in different manners.

In a sense, the option (i) corresponds to the negative, and option (ii) to the affirmative answer to the question "Could I have been born a Martian?"[3] Since the range of answers to this central question is thus exhausted, it seems that we have reasons to reject NOR in this particular counterfactual world. However, even in the case of accepting option (ii), it is quite legitimate to add the information on whether you are an Earthling or a Martian to the information of one's own existence in a particular point in history of intelligent beings native to the Solar system. This is something which lies outside of the scope of the conventional DA, but presents its generalization rather than a special case. Now, it looks plausible to assume that this additional information should actually strengthen one's belief in the impending doom. It is perhaps easiest to see why when you think of the very meaning of "doom" in such circumstances. In contradistinction to the "single intelligent race" case, now the doomslike consequences encompass both the general case of destroying intelligent beings in the entire Solar system (due to the explosion of a nearby γ-ray burst, for instance), and the particular case of destruction of such beings only on **human** home planet (due to a collision with an asteroid, for instance).

In total, the existence of such a declining Martian population would rather have strengthened our belief in impeding doom–something completely contrary to the spirit of NOR. Now, we may return from that counterfactual world to the actual one, in which there are no Martians (or at least no multicelled, intelligent ones), and try to perceive the impact of this absence on our probabilistic calculations. It seems utterly unreasonable to ascribe to possible extraterrestrials living parsecs or kiloparsecs away what we have denied our neighbours, the Martians in the example above. If the distance plays no significant role, it seems that we are left with very few reasons to accept NOR.

4. POSSIBLE SOURCES OF NOR

In this Section we analyze possible sources of credence in NOR. We shall see that these are either inconclusive or irrelevant, or just reflections of wishful thinking. Of course, here we exclude the basic such source or motivation, that is, the desire to avoid accepting at face value the unpleasant conclusions of DA. We have seen, in the previous Section, that the *bona fide* attempt to overcome DA with NOR is highly counterintuitive, and based upon very dubious premises. If NOR is to be taken seriously, some independent support has to be found; here we investigate possible sources of such support.

4.1. Computationalism

The essence of this widespread opinion is that cognition (as the major aspect of intelligent observers) is computation, along with sufficiently complicated algorithm (for a review, see for instance Shapiro 1995). It is usually claimed to be only a "working hypothesis", but in practice it is standardly taken much more seriously, especially in circles of cognitive and computer scientists, and by some physicists and cosmologists as well (e.g. Tipler 1994; Deutsch 1997).

Now, computationalism is related to NOR, since it is commonly regarded as an **explanation** of why all intelligent observers should be included in the reference class. This is particularly strong argument in various forms of transhumanist ideas (e.g. Drexler 1987; Moravec 1988). However, it does seem to beg to be questioned, since it presupposes a property of objects never empirically investigated, and on top of that asserts that the same property (not proven to be coherent or even sensible, at that!) enables us to classify these objects and put them in the same reference class. There does seem to be a strong circular element in this way of reasoning.

In other words, the application of computationalism to the anthropic selection effect problem is doubly dubious. First of all, human minds may not be the computable after all (as many quite respectable researchers have suggested; e.g. Penrose 1989), and it seems that **this** explanatory task has clear priority over the solving of anthropic puzzles like DA. Secondly, computationalist dogma indicates that a construction of true, human-level or higher artificial intelligence (AI) is a matter of near future.[4]

[2] This is certainly so if we suppose that (1) Martian population is, like terrestrial, finite overall to avoid problems with infinities (we shall return to this later), and (2) Martian population in the past has not been so high as to cause what might be called an "inverse" Doomsday Argument, that is our posterior probability of living thus late in the overall history of intelligent beings in the Solar system is diminished.

[3] Note that the true answer to this question crucially hinges upon the definition of "I", which is one of the most long-standing and difficult problems in the entire history of philosophy. We cannot enter into this dispute here.

[4] Parenthetically, let us recall that the father of the modern computers, Alan Turing, predicted that we were to have thinking machines well before the year 2000!

Future AIs could, arguably, play the same role in a generalized form of NOR as aliens in Bostrom's version quoted above. The fact that AIs come later in history than humans is irrelevant, since some aliens could as well be younger than ourselves somewhere in the Galaxy. On the other hand, is it not an inclusion of AIs in the reference class in the DA reasoning manifestly absurd? If AIs entirely replace humans at some point in the future, this would in fact **prove** the validity of DA-based prediction, not falsify it. If humanity persists in parallel with AI, should we not be normally applying Bayesian reasoning (with possibly slightly modified priors) all along?

In brief, it seems somewhat ironic that Bostrom, correctly, criticizes Korb and Oliver in stating that (Bostrom 2002)

> Korb and Oliver's application of the Doomsday argument form to individual life spans presupposes a specific solution to the problem of the reference class. This is the problem, remember, of determining what class of entities from which one should consider oneself a random sample...

But at the same time NOR does something similar: presupposes, in a wide and rather vague way, the properties of an all-inclusive reference class.

4.2. Carter's Argument

The argument due to astrophysicist Brandon Carter (godfather of all "anthropic principles") purports to indicate extreme scarcity of life and intelligence in the universe. This argument has usually been taken very seriously by proponents of the anthropic principles and anthropic thinking in general (e.g. Barrow and Tipler 1986; Bostrom 2002a; Tipler 1994). Since it is not widely known, especially among physicists and astronomers, we shall summarize it here.

If characteristic astrophysical (τ_*–say the stellar evolution) and biological (τ_l–say the eucaryotes evolution) timescales are truly uncorrelated, life in general, and intelligent life in particular, forms at random epochs with respect to the characteristic timescales of its astrophysical environment (notably, the main-sequence lifetime of the considered star). In the Solar system, $\tau_* \approx \tau_l$, within the factor of two. However, in general, it should be either $\tau_l >> \tau_*$ or $\tau_* >> \tau_l$. In the latter case, however, it is difficult to understand why the very first inhabited planetary system (that is, the Solar System) exhibits $\tau_* \approx \tau_l$ behaviour, since we would then expect that life (and intelligence) arose on Earth, and probably at other places in the Solar System, much earlier than they in fact did. This gives us probabilistic reason to believe that $\tau_l >> \tau_*$ (in which case the anthropic selection effects explains very well why we do perceive the $\tau_* \approx \tau_l$ case in the Solar System). Thus, the extraterrestrial life and intelligence have to be very rare, which is the reason why we have not observed them yet.

If the number of aliens within our cosmological horizon iz zero or very small, it will not influence any statistical considerations, and DA will be as valid (or invalid) as it was apart from considering any aliens. However, as we shall discuss later, the mention of horizons (and their heterogeneous nature) is not made anywhere in the statement of NOR. This leads to a strange situation that, since the universe is most probably open, containing infinite number of galaxies, even extremely small chance of emergence of life and intelligence will lead to the infinite number of alien beings and civilizations. Thus, Carter's argument is essentially ineffective as a remedy to the DA problem. However, it does give a sort of psychological advantage to proponents of NOR, who effectively postpone the neccesity to deal with any alien observers till some far, far future (see, for instance, Bostrom 2002b).

This said, it's worth mentioning that Carter's argument is inherently on rather shaky legs. Several criticisms of it appeared in the literature, two most important being those of Wilson (1994) and Livio (1999). There is a significant interest in deepening their criticisms by including other possible sorts of correlations between the two timescales, thus undermining the crucial assumption of the argument (Ćirković and Dragićević, in preparation).

4.3. "Finitism"

The realization that there are most probably infinitely many galaxies in the universe, slowly diffused from the narrow circle of theoretical cosmologists to the wider, including philosophical, audience. In particular the philosophical community (including referees!) shows strong resistance to the idea of infinite universe implied by Einstein's general relativity for density smaller than or equal to the critical density. Infinity of the universes bears many strange conceptual consequences, some of them pointed in the study of Ellis and Brundrit (1979). But, there is no indication that these consequences (like existence of exact duplicates of all humans that ever existed) are leading into incoherencies or contradictions.

Fortunately, we are spared the empirical verification of these bizarre features by the presence of cosmological horizons (e.g. Ellis and Rothman 1993). However, there are two possible kinds of horizons in cosmology: particle horizons and event horizons. All models ever thought to be realistic, including the standard Big Bang cosmologies, possess particle horizons; roughly speaking, it is equivalent to saying that light from objects more distant than some value simply did not have time to reach us since the Big Bang. If this were the only type of horizon in the real world, we would have a chance to eventually meet our doubles and empirically verify some of the strange features of an infinite universe at some future epoch. However, the things are not so simple. According to the new results confirming existence of a large positive cosmological constant, there are not only particle horizons, but the event horizons as well, i.e. those surfaces which prevent communication between two co-moving observers at **all** times (e.g. Ćirković and Bostrom 2000).

The formulation of NOR given by Bostrom contains no mention of horizons of any type. Thus, it is immediately liable to the criticism that it enables a form of causality violation in that some (albeit only statistical) influence propagates instantaneously (see also Olum 2002).

4.4. "Galactic Club" and liberalism

In our opinion, a further impetus for NOR may be obtained from the early SETI optimism of 1960s, carried by ideas of Drake, Sagan, Shklovskii, Bracewell, and others. As is well-known, the enthusiasm receeded in the meantime, but some of its main ideas deeply underlie the contemporary philosophical and scientific thought and even popular culture. We find Tipler's (1981, esp. Appendix II) criticism of some of the motivations behind SETI highly justified.[5] NOR is the other side of the coin of the same desire for immortality, known since Gilgamesh, which Tipler finds (and rightly ridicules) in Drake and some other contact-optimists. By assuming the maximally comprehensible reference class, NOR instills the idea that whenever and wherever we encounter extraterrestrials, they will be able to solve or help solving our own problems. Moreover, it suggests that even their very existence **anywhere in the universe** is, in a sense, solution to the DA problem; and although people may profess to disbelieve the existence of ETIs, on a psychological level the impression of vastness of the universe certainly makes one less worry about DA if one accepts NOR. This solution is unsound, simply speaking, since it prejudicates the solution to the reference-class problem, the most important **open** problem in the field of anthropic reasoning.

Really problematic aspect of this attitude consists in its profound political consequences here on Earth. It's clear that there are people, in liberal and leftist circles, who would certainly find such ideas about "cosmic club" or even "cosmic brotherhood" a natural extension of the fight for freedom and universality of human rights among human observers on this planet. However, the extension is more than dubitable, and leads to a sort of anthropocentric caricature, nonsensical in the same manner as its opposite picture of "alien invaders" or "buggy-eyed monsters", well-known from (bad) SF movies and literature. Attempting to "understand" possible alien beings as "human beings with strange physical shape and/or skin color" are in reality attempts to both ignore the problem and to insult both reason and imagination, since the tacit assumption here is that (i) to be a "human being" is something universally exalted by definition, and (ii) we cannot really comprehend something or somebody that is profoundly different from ourselves. Ironically and indicatively enough, the attitude (i) has been most loudly proclaimed in XX century by communist ideologues, like Maxim Gorky, i.e. intellectual collaborators in the greatest mass murder in human history, and indescribable sufferings of hundreds of million of those same human beings.

Among other things, it is said from time to time that AI studies will enable us to investigate an "alien" (in the limited sense of non-human) intelligent being soon enough. As such, the idea permeates literature in transhumanism, futurology and even cosmology (paradigmatical example of the latter is Tipler's Omega-point theory). We find it to be another instance of rather wishful thinking on behalf of people proposing optimistic, melioristic views of the universe.

4.5 Self-Indication Assumption

Finally, there is a sensitive issue of the so-called Self-Indication Assumption (henceforth SIA). Roughly speaking, the idea is that an observer is more likely to observe anything (i.e. to find him/herself alive) if there is a large collection of observers compared to the case of a small collection of observers. This assumption has, surprisingly, not been precisely defined in the papers proposing or defending it, but Bostrom (2000)–who criticized it–proposed a name Self-Indication Assumption, as well as the definition we may use for a start:

> **SIA:** Given the fact that you exist, you should (other things equal) favor hypotheses according to which many observers exist over hypotheses on which few observers exist.

Other locutions used in the literature (e.g. "one is more likely to find oneself in the long-lived race", Olum 2002) are equivalent to this. The fact that SIA **exactly** compensates for the DA-inducing probability shift has been demonstrated by Kopf et al. (1994), and it remains beyond doubt. However, the other merits and demerits of SIA have remained an open question; it has been criticised by Leslie (1996) and Bostrom (1999, 2000).

Again, ironically enough (the subject seems permeated with irony), the thought experiment Bostrom proposes as a justification for NOR–the one with a million "small" and a million "large" civilizations in the universe–is actually invoked by the proponents of SIA, notably Olum (2002). The fact that SIA cancels the DA has been known since the work of Kopf et al. (1994), but the possibility that something rather vague like NOR can cancel it, certainly sounds presumptuous!

5. INFINITY OF OBSERVERS?

In an open universe, the total number of intelligent observers is infinite; in the closed case, it is necessarily finite. How reasonable is, then, that a

[5] Although, of course, we do not endorse other features of Tipler's view, notably his reliance on Carter's argument (see above), or his pseudo-theological conclusions.

tiny change in the cosmological density fraction $\Omega = 1 \pm \varepsilon$, which represents the boundary between the two cases (open *vs.* closed universe) and which manifests itself on the timescales of $\sim 10^{10}$ years, may cause the drastic change in the epistemic probability of us being extinct on the timescale of $\sim 10^1$ years? Not very much, arguably.

If humankind is not entirely miraculous, there is a finite probability density for a technological civilization to arise at each point of spacetime. By virtue of the cosmological principle (large-scale homogeneity and isotropy of matter and spacetime) all these probability densities, when averaged over a large enough volume, must be the same. In an infinite universe, this immediately means that the number of technological civilizations is infinite. However, the infinite number of civilizations entails multiple copies of every civilization, including multiple copies of ourselves as observers (Ellis and Brundrit 1979). Thus, **all relevant** reference classes are infinite, and all auxiliary anthropic assumptions become vacuous: in the same sense as it seems fallacious to derive probability shifts from any finite rank in an infinite set of observers (DA), it seems that there is no predictive power in SIA if the set of observers is (regardless of the actual theory used) infinite. How do we compare various probabilistic theories if we know that there is an actual infinity of observers in any case? Bostrom (2002a) correctly notes that infinities cause many problems in this area of statistical reasoning (Pascal's Wager, etc.), but he leaves no prescription what to do when cosmology **forces** such a situation on us.

(And even if we do not trust the cosmological principle, it is enough to accept much weaker assumption that variations between spatial regions are not too large to prevent intelligence from arising in entire, but infinitesimal fraction of all spacetime. Since intelligent communities/civilizations are discrete entities, the relevant densities cannot asymptotically approach zero, and therefore any sum over infinitely many regions must diverge.)

6. DA vs. SETI: FALLACITY OF GOTT'S ARGUMENT

Obviously, the validity (or otherwise) of NOR bears great relevance to one of the most interesting projects ever conceived in the history of humankind: the search for extraterretrial intelligence. In any consideration regarding the probable evolution and fate of intelligent observers and their communities, we are necessarily limited by the absence of other such communities known to us. This circumstance, together with our knowledge of the relatively large size both in space and in time of (at least in principle) inhabitable universe, leads to the problem of "Great Silence" (Brin 1983), sometimes dubbed the Fermi's "paradox" (although it is certainly misleading to speak of it as a paradox in a methodological sense) or "the astrosociological problem" (Kardashev and Strelnitskij 1988; Lipunov 1997). Namely, we expect to perceive traces of activities of advanced civilizations in our past light cone, since the age of the Galaxy is significantly larger than the age of Earth and the Solar system.[6] Since we have not perceived such traces, this puts a strong empirical constraint on the character and distribution of intelligent communities in the Galaxy.

Proponents of DA sometimes claim that it offers a solution to the puzzle. Gott (1996) thus writes:

> Another point... is that you find yourselves born on the home planet where your species originated. If that is not special, then a reasonable fraction of all intelligent observers must still be on their home planets. This answers Fermi's question: Where are they? (Answer: a significant fraction are still at home just like us.) This is sufficient to explain why we have not been colonized.

Similar to Gott's idea is the one hinted at by Leslie, but immediately rejected by himself (Leslie 1996, p. 192):

> Why... do we see no sings of extraterrestrial intelligent beings? Might there be very many technologically advanced civilizations in space-time as a whole, but only very few, and those ones unusually small, at the early time at which I am living? Just conceivably this scenario is correct. But an observer in a technological civilization would be far less likely to be in an early period, a period when such civilizations were small and few, than in a later one when they were huge and many. I therefore have grounds for thinking the scenario wrong. This could mean that my technological civilization wasn't among the very earliest: many others had developed previously without making their presence known to us humans, probably because they quickly became extinct. Alternatively it could mean that only a very few technological civilizations will ever have developed, in the entire history of the universe.

In our oppinion, Gott's argument (and Lesli's hint) is fallacious for the following reasons. First of all, it actually leaves out the key ingredient of DA, that is the Doomsday itself–the reasoning which applies to humanity should apply to other intelligent species as well (it is not so obvious, and therefore

[6]This conclusion holds even if we consider the "effective" age of the Galaxy defined as the period of time for which the Galactic chemical evolution maintained conditions favorable to appearance of life. As discussed recently by Livio (1999), such conditions arose only a billion years or so after the epoch of galaxy formation. Therefore, we conclude that an effective age is still $\sim 10^{10}$ yrs. By comparison, estimates of the timescale necessary for colonizing the entire galaxy through self-replicating von Neumann probes is, depending on the exact model, between 10^7 and 10^8 yrs.

NOR has been formulated by Bostrom and others; more on this below). It is improbable (although such contrived scenarios are imaginable) that the population of a particular advanced intelligent species decreases **in the same time** as that species expands and investigates many planetary systems in the galaxy. It is reasonable to conclude that most of species simply do not reach the stage at which the division between the population on the home planet and colonizing population is significant.

In addition, a simple analogy with the colonization of Earth's surface shows the fallacity of this argument. Let us imagine an observer living in pre-Columbian America or on some Polynesian island, and consider his opinions about civilizations arising on different parts of the Earth's surface (effectively "other worlds"). His analogue of the Fermi's paradox will be as well-defined as our deliberations on the SETI questions. Notably, he might conclude that other landmasses are uninhabited, since he perceives no traces of such visitors, and since there had been enough time for advanced societies to arise there. We know, a posteriori, that different civilizations **did arise** on various parts of Earth's landmass, and that some of them eventually reached all inhabited parts of our planet. But in each case, by far the largest part of the population of colonizing civilizations simply stayed at home (if at all aware of the colonizing endevoar!). *Per analogiam*, lacking any external knowledge whatsoever about the motivations and capacities of advanced extraterrestrial civilizations, we may judge that the case of human expansion is a typical one, and therefore invoke the very Copernican assumption to undermine Gott's "explanation" of Fermi's paradox.

The problem with such an explanation, in addition, is that it *in principle* violates some basic ideas about the nature of systems sufficiently complex to be called intelligent. In agreement with the weak reductionist assumption that all systems, including intelligent and conscious ones, are subject to the laws of thermodynamics, this means that all observers are bound to consume free energy and produce entropy for any, even the simplest acts of consciousness, like the computation.[7] In principle, a sufficiently uniform emergence of conscious communities throughout the universe would result in the existence of such communities of age comparable to that of the universe itself. This seems strongly precluded by our astronomical knowledge, as Fermi and other have noted, even without the actual physical presence of extraterrestrials in our midst. Conceptual nature of this argument is emphasized since its proponents can always claim that this uniform rate is very low, such that its characteristic timescale is still large compared to the present age of the universe (and the structure inside it), and thus the emergence and disappearance of conscious communities did not have enough time to reach an equilibrium state. In other words, the Copernican assumption about emergence of life can yet be saved, but at a large price: it is claimed that intelligent life is "almost" miraculous, since for various reasons the preconditions for it are only very rarely satisfied (e.g. Ward and Brownlee 2000).

However, this chain of reasoning still requires a particular well-defined ingredient in the history of each civilization: its disapperance from the scene. As one may note, if the basic conclusion of DA is correct, this may as well be solution to the Fermi's paradox *a fortiori*. Among various apocalyptic scenarios considered in some detail in the book of Leslie (1996), it is worth noticing that even some of the non-catastrophic ones, like the "transcedence" of any kind, possibly resulting in a collective consciousness, would also act to reduce the efficient cross-section for interaction of the civilization with external cosmos, therefore producing the empirically established "Great Silence" (e.g. Brin 1983). This circumstantial argument does lend some support to serious study of DA and its possible physical realizations. One thing seems rather certain: the danger of extinction is the heaviest at some early points in the history of each intelligent community, since once this community has learnt to use resources beyond its home planet, and possibly beyond its home planetary system, the risks become considerably smaller, even if still non-zero.[8] This is the picture of "Great Filter" of Hanson (1998) (http://hanson.gmu.edu/greatfilter.html). DA agrees with the sharp nature of this filter.

But is DA effective on wider, galactic scale at all? As we have seen, Bostrom (2000) claims that **DA should always apply only to the entire reference class**. In his interpretation, it effectively rules out DA if there are extraterrestrial intelligent beings of plausible characteristics (more precisely, approximately uniform distribution of longevities of their societies). This requirement has been previously argued against by Leslie (1996), and recently by Olum (2002). Especially significant is Olum's criticism which is based on physical causality, stating that it is senseless to claim that beings which exist outside of our causally connected region (and in Bostrom's version, most of extraterrestrials would actually be located outside of our particle horizon) could change our probability priors. Olum's study purports to support SIA, stating that one's exis-

[7]While, of course, we do not claim that computation is sufficient condition for intelligence (as proponent of the so-called strong AI conjecture would), but it is reasonable to have it as a necessary condition. This leaves open the option recently strongly defended by Penrose (1989, 1994), that consciousness necessarily includes non-computable elements.

[8]Possibility of truly cosmic catastrophes, like the decay of false vacuum (if we are unhappy enough to live in one) or tearing apart the fabric of spacetime due to emergence of naked singularities (Barrow and Tipler 1986), will still present some risk for a community of any age. Still, however, even the most pessimistic estimates give such dangers orders of magnitude smaller prior probabilities than more local challenges, like the nuclear war, extinction due to ecological disruptions or asteroidal/cometary impacts. The latter will be considered obsolete (as the sources of extinction and not, of course, as great dangers to be avoided!) once first self-sufficient extra-planetary communities are established, in human case when we establish colonies on Moon or Mars, or even in the Earth's orbit (Bostrom 2002b).

tence *a priori* indicates existence of a large number of observers. Although Bostrom explicitly rejects SIA (which is indeed based on very suspicious–to say the least–argumentation), it is interesting and indicative that in justifying the no-outsider requirement he uses exactly the SIA-like argument:

> The Doomsday argument... only served to cancel the effect... that you were more likely to turn out to be in the human species given that the human species is one of the large rather than the one of the small civilizations. This shows that if we assume we know that there are both "large" and "small" extraterrestrial civilizations–the precise numbers in the above example don't matter–then the right probabilities are the ones given by the naive empirical prior. Only if there are no "outsiders" (extraterrestrial civilizations) does DA work as intended.

To the present authors, this argumentation seems rather poorly founded. Apart from correct Olum's criticism, one may always ask what independent reason one has to put this or that extraterrestrial civilization in the reference class. This does not seem like something that can be established *a priori.* Rather, except explicitly metaphysical assumption that intelligence is somehow so strictly specified that one cannot make a misidentification of it (very bad form of epistemic anthropocentrism), there seems to be no clear physical reason for such a belief.[9] After all, if it is so easy to solve the DA problem in this manner, why not just admit that there are other intelligent beings on Earth, like dolphins and whales, and we have changed our empirical prior, and everybody is happy! In addition, the idea that the distribution of longevities will be so uniform as in the Bostrom's example (million "small" and million "large" civilizations) flatly contradicts astrophysical realities of any civilization's environment. For instance, one may claim that resources necessary for any form of information processing (characterizing any intelligent being and any community of such beings) are "quantized", since the natural steps in development of any society include its home planet, home planetary system and a number of surrounding planetary systems, its home galaxy, galaxy group, etc. The risks of extinction are expected to vary in inverse proportion to the spatial volume "technologized" by a community, and therefore the relative increase in probability of surviving past some fixed time t by acquiring (say) resources of a single planet is very different for "small" and "large" civilizations (and any intermediate case as well).

The underlying idea we would like to suggest is that DA **without the no-outsider requirement** offers a consistency argument for the most reasonable solution of the Fermi's paradox. Such solution suggests that **intelligent species are generically truly frequent in the universe, but the number of those able to survive the "Great Filter" is very, very small.** In other words, our *a priori* confidence in the individualistic interpretation of Final Anthropic Hypothesis (FAH)[10] is significantly decreased when we are faced with both DA and Fermi's paradox. This problem does not degenerate in a vicious circle in the realistic case–as occurs in the steady-state cosmological model, for instance–because the age of the universe is finite, and therefore, it is plausible that the technologization, which behaves essentially as diffusion processes in statistical physics, has not yet reached the "equilibrium" state.

DA, as Leslie has repeatedly emphasized, is not something "mystical" or "fateful", but just a reason for rational reassessment of probabilities of extinction through physical and specific (albeit not necessarily known at present) causes. Let us suppose for a moment that the true average risk of extinction of a young (i.e. not-yet-space-faring) civilization in the Galaxy through cometary/asteroidal bombardment is thousand times larger than our empirical assessment on the basis of Earth's past in the Solar system suggests. In that case, although still not doubting natural causes of our existence, we may be rather certain that the number of other societies in the Galaxy is very small, so small indeed that we may well be the only one galactic society on the verge of serious space travelling capacity. This is in accordance with the Fermi's observation, and also in agreement with the DA conclusion as far as the other intelligent species share some common properties with us (basically the nature of their intelligence is such that the theory of observational selection effects applies *mutatis mutandis*). The anthropic selection effect obviates the need for answering the question why is it exactly **us** (i.e. *homo sapiens*) who are "chosen" in this manner.

[9] This is not to say that such reason is unimaginable or even unreasonable. This oecumenical assumption can be supported, for instance, by establishing the clear link between the properties of intelligence and consciousness and cosmological initial conditions fixed at the Big Bang singularity (or appropriately generalized spacetime boundary). In fact, one of the present authors (M. M. Ć.) has recently proposed, with collaborators, exactly one similar scheme, based on the possible resolution of the quantum measurement puzzle (Dugić, Ćirković and Raković 2002). However, any such idea has still a long way to go before being accepted on physical, rather than metaphysical grounds.

[10] We refer to the following conjecture: *Once intelligent information processing comes into existence, it will never die out* (Ćirković and Bostrom 2000). FAH is susceptible to various differing interpretations. In particular, we need to distinguish between the following two meanings: (1) There is at least one intelligent race in the universe that will continue to exist indefinitely. For the sake of brevity, we shall call this interpretation *individualistic.* (2) Any particular intelligent race might eventually die out, but intelligent life as a whole will exist indefinitely. This interpretation may be termed *holistic.* Statement 1. logically implies statement 2., but not *vice versa.*

7. CONCLUSIONS: NO SALVATION FROM LITTLE GREEN MEN

The interplay between cosmology and philosophy has–since Keppler and Galileo–often showed new and in many cases unexpected insights. However, this success story should not lull us into a belief that it is easy to solve important problems by simply sailing on–sometimes murky–waters of philosophical arguments. This is the case with the No-Outsider Requirement.

It is worth noticing that recently Bostrom has partially revised his stance on NOR and DA. Now, his claim is (Bostrom 2002a; several private communications) that **if** we had good direct evidence of the statistical distribution of the longevity of extraterrestrial civilizations (e.g. if we **knew** that half of them were long-lasting and half of them short-lived, as in the example quoted above) then this would diffuse the DA. This presents a significant and bold move to avoid the most serious of the problems facing NOR as described in this essay. In Bostrom's current view, seemingly, NOR teaches that one cannot in general arbitrarily choose to omit observers from one's reference class. One should note that this, however, requires an additional information (the one which is very difficult to come by, ultimately relying on astrobiology) in order to judge whether we should, after all, heed the doomsayer's warnings or not.

The claim that NOR should hold in the particular case of DA as exposed by Gott and Leslie, bears some emotional charge as well, since it is the same feeling of cosmic loneliness which frightened Pascal and in modern times motivated both serious SETI (and, one suspects, AI as well) efforts and various UFO cults and enterprises. However, this should be spelled out explicitly: the thought that aliens can, even by their very presence, save us from doomsday (which, as both Leslie's book and the recent study by Bostrom (2002b) indicate, is in most scenarios a consequence of our own actions) tells us more about 'growing pains' of humanity's cosmic infancy than about the universe, its inhabitants or its probabilistic aspects.

In one very specific sense NOR deserves a justification, but it is ethical, not probabilistic. If one day our SETI efforts are crowned with success and we discover an advanced extraterrestrial society (or at least incontrovertible proofs of its existence), we shall obtain the final proof that life and intelligence are not flukes, but a ubiqutous and persistent phenomena of nature. This, in turn, might bolster our self-confidence in the face of great challenges of the future, like colonization of space or building a truly stable and humane society. If–as seems quite plausible–the greatest source of doomsday risks are human actions, then this situation might alleviate some of the perils. But this understanding of NOR seems remote from its original proposed role in Bayesian reasoning.

To summarize, the reference class objection of Korb and Oliver should be taken as a serious possibility of correct relativization of the reference class (and thus offers a ray of hope in solving of this difficult problem in the field of anthropic reasoning). Before the situation in respect to the reference class problem is investigated in more detail (which certainly lies beyond the scope of the present study), it is more reasonable to conclude that any extraterrestrial intelligent observers do not properly belong to the reference class of relevance for DA. Accordingly, we conclude that arguments against DA should be sought on different sides, and that aliens cannot–apart from the case of actual contact and physical interaction–influence either our fate, or our calculations about it.

Acknowledgements – It is a pleasure to thank Nick Bostrom for great inspiration and kind support. One of the authors (MMĆ) has been partially supported by the Ministry of Science, Technology and Development of Serbia through the projects No. 1196, "Astrophysical Spectroscopy of Extragalactic Objects" and No. 1468, "Structure and Kinematics of the Milky Way."

REFERENCES

Adams, F.C. and Laughlin, G.: 1997, *Rev. Mod. Phys.* **69**, 337.
Adams, F.C. and Laughlin, G.: 1999, The Five Ages of the Universe (The Free Press, New York).
Barrow, J.D. and Tipler, F.J.: 1986, The Anthropic Cosmological Principle (Oxford University Press, New York).
Bell, J.S.: 1987, Speakable and unspeakable in quantum mechanics (Cambridge University Press, Cambridge).
Bostrom, N.: 1999, *Mind* **108**, 539.
Bostrom, N.: 2000, Observational Selection Effects and Probability (PhD thesis, London School of Economics).
Bostrom, N.: 2001, *Synthese* **127**, 359.
Bostrom, N.: 2002a, Anthropic Bias: Observation Selection Effects (Routledge, New York).
Bostrom, N.: 2002b, *Journal of Evolution and Technology* (online journal at http://www.jetpress.org/index.html), vol. **9**.
Bostrom, N. and Ćirković, M.M.: 2003, *Phil. Quart.* **53**, 83-91.
Brin, G.D.: 1983, *Q. Jl. R. Astr. Soc.* **24**, 283.
Ćirković, M.M. and Bostrom, N.: 2000, *Astrophys. Space Sci.* **274**, 675 (preprint gr-qc/ 9906042).
Deutsch, D.: 1997, The Fabric of Reality (Penguin Books, New York).
Drexler, K.E.: 1987, Engines of Creation (Anchor, New York).
Dugić, M., Ćirković, M.M. and Raković, D.: 2002, *Open Systems and Information Dynamics* **9**, 153.
Ellis, G.F.R. and Brundrit, G.B.: 1979, *Q. Jl. R. Astr. Soc.* **20**, 37.
Ellis, G.F.R. and Rothman, T.: 1993, *Am. J. Phys.* **61**, 883.

Gibbons, G.W. and Hawking, S.W.: 1977, *Phys. Rev. D* **15**, 2738.
Gott, J.R.: 1993, *Nature*, **363**, 315.
Gott, J.R.: 1994, *Nature*, **368**, 108.
Gott, J.R. III: 1996, in Clusters, Lensing and the Future of the Universe, ed. by V. Trimble and A. Reisenegger (ASP, San Francisco), p. 140.
Hanson, R.: 1998, preprint at http://hanson.berkeley.edu/greatfilter.html.
Kardashev, N.S. and Strelnitskij, V.S.: 1988, in Bioastronomy – The Next Steps, ed. by Marx, G. (Kluwer, Dordrecht), p. 295.
Kopf, T., Krtous, P. and Page, D.N.: 1994. preprint gr-gc/9407002.
Leslie, J.: 1992, *Phil. Quat.* **42**, 85.
Leslie, J.: 1993, *Mind*, **102**, 489.
Leslie, J.: 1996, The End of the World: The Ethics and Science of Human Extinction (Routledge. London).
Lewis, D.: 1986, Philosophical Papers (Oxford University Press, New York).
Lipunov, V.M.: 1997, *Astrophys. Space Sci.* **252**, 73.
Livio, M.: 1999, *Astrophys. J.* **511**, 429.
Moravec, H.P.: 1988, Mind Children: The Future of Robot and Human Intelligence (Harvard University Press, Cambridge).
Olum, K.: 2002, *Phil. Quart.* **52**, 164 (preprint gr-qc/0009081).
Oppy G.: 2001, *Philo*, **4**, no. 2 (online journal at http://www.philoonline.org/).
Penrose, R.: 1989, The Emperor's New Mind, (Oxford University Press, Oxford).
Penrose, R.: 1994, Shadows of the Mind. A Search for the Missing Science of Consciousness (Oxford University Press, Oxford, 1994).
Shapiro, S.C.: 1995, *Minds and Machines*, **5**, 517.
Tipler, F.J.: 1981, *Q. Jl. R. Astr. Soc.* **22**, 279.
Tipler, F.J.: 1986, *Int. J. Theor. Phys.* **25**, 617.
Tipler, F.J.: 1994, The Physics of Immortality (Doubleday, New York).
Tipler, F.J.: 1999, *Astrophys. J.* **511**, 546.
Ward, P.D. and Brownlee, D.: 2000, Rare Earth: Why Complex Life is Uncommon in the Universe (Springer-Verlag, New York).
Wilson, P.A.: 1994, *Brit. J Phil. Sci.* **45**, 241.

ВАНЗЕМАЉСКА ИНТЕЛИГЕНЦИЈА И СУДЊИ ДАН: КРИТИЧКО ИСПИТИВАЊЕ ЗАХТЕВА ЗА ОДСУСТВОМ ТУЂИНАЦА

М. М. Ћирковић[1] и В. Милошевић-Здјелар[2]

[1]*Астрономска опсерваторија, Волгина 7, 11160 Београд 74, Србија и Црна Гора*

[2]*Physics Department, University of Winnipeg, Winnipeg MB, R3B 2E9, Canada*

UDK 52–37
Оригинални научни рад

Такозвани "Аргумент Судњег дана" привукао је доста пажње у научним и филозофским круговима последњих година. Један од начина предложених да се неприхватљив закључак овог аргумента неутралише јесте тзв. "захтев за одсуством туђинаца" који сугерише да присуство других интелигентних посматрача ("туђинаца") обесмишљава резоновање које води ка апокалиптичком закључку. Очигледно, ова аргументација је релевантна не само за ефикасност аргумента Судњег дана, већ и за питања као што су референтна класа у антропичком расуђивању, трансхуманизам и теорија SETI пројеката. У овом раду ми критички преиспитујемо захтев за одсуством туђинаца и закључујемо да је он или погрешан или ирелевантан.