

BOOK REVIEW

Nick Bostrom, *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. New York & London, USA & UK: Routledge, 2002. xiii + 224 pp., \$70 (hardcover). ISBN 0-415-93858-9.

The centuries-old question *why the universe is just so?* has only recently begun to be discussed in serious quantitative manner. Above all, this is due to the vast accumulation of physical and astronomical knowledge in the course of the twentieth century. With the recent discoveries of the geometrical flatness of the universe at large, a greater than previously suspected cosmological vacuum energy, as well as planetary systems around dozens of nearby stars, we are in good position to try to reassess humanity's position in the universe and reevaluate the lessons of the Copernican revolution. What modern cosmologists have begun to discuss in earnest is the link between our own existence as intelligent observers evolved on a typical planet from the simplest procaryote lifeforms over billions of years and the properties of universe (and other universes!) at largest scale. This link is technically called an observational selection effect, although in both research and popular literature it is often known under the slightly misleading title of anthropic principle(s). If from now on anybody wishes to seriously study these matters, *Anthropic Bias* is without question an excellent place to start. Rapid developments in cosmology, astrobiology, quantum theory (and quantum technology!), as well as in philosophical discussions on the foundations of these scientific disciplines, make a book like this desirable for scientists and philosophers of diverse interests, indeed.

And Bostrom's book performs its task splendidly enough. It makes an amusing, although at times quite exacting, reading. Bostrom is a witty and refined author, whose text flows smoothly,



and whose pertinent digressions have something of the charm of writings of Feynman or Popper. From quantum cosmology to annoying traffic jams, from quantum mechanics to Adam and Eve thought experiments, from freak observers created by black hole radiation to the (in)famous Doomsday argument of Gott, Carter and Leslie (not to mention future totalitarian world government and the Principal Principle of the late David Lewis), the book reads like an exciting novel. It first exposes the deep mystery of the anthropic “coincidences” or “fine-tunings”, and then, like a true detective, tries to gather many clues from various fields of science and philosophy in order to solve the case. At the end, it presents us with a sort of sketchy reconstruction of the way the mystery has arisen in the first place.

Thus, there are three basic, tightly interrelated parts of the exposition. The first (roughly the first three chapters) deals with the “classical” anthropic thinking, definitions of anthropic principles, and examples of “fine tuning” revealed in the numerical values of constants of nature and cosmological parameters, from Eddington and Dirac to this day. Bostrom attempts to bring some order into the vast jungle of confusion as to the definitions and meanings of various “anthropic principles” (he counts more than 30 of them in the literature). He succeeds splendidly in this difficult task, and allows the reader to gain a completely new perspective, which is obscured in some of the most authoritative works on the subject thus far.¹

This part is particularly useful as an antidote to the misconstrual of the anthropic thinking – still prevalent in some circles, slightly more in philosophy of science than in physics and cosmology – as anthropocentric or “cozy” or teleological. This line of “argument” goes like this: anthropic principles point to peculiar properties required by the (part of the) universe we inhabit. Since these are of measure zero in the set of all possible properties (or values of physical and cosmological parameters), and since we obviously exist, our domain has been intentionally created (or “fine-tuned”) for the purpose of our existence. All sorts of teleological and theological inferences have been drawn from that point onward. But Bostrom clearly and decisively demonstrates that this is not the case; the line of reasoning above is a misconstrual. The prevalence

of the the confusion and mispresentation of anthropic reasoning is attested to by the fact that both some theologically-minded *and* some materiastically-minded philosophers and scientists have insisted on this modern version of the classical design argument. Yet, as Bostrom amply shows, anthropic reasoning may and should be interpreted not just disteleologically, but rather *anti-teleologically*. That is, instead of pointing out something really special about our cosmological domain and its physics (and, by extension, ourselves), it emphasizes that any special feature we might observe is nothing but an illusion, a necessary consequence of our restricted viewpoint. Since we could not exist in other places (for instance in those in which there is no resonance in ^{12}C nucleus, enabling formation of elements heavier than helium), we will not observe these places, no matter how real and common they are. Here, Bostrom highlights the central aspect of anthropic thinking: the *anthropic principle as expressing an observational selection effect*. Physicists and astronomers have been familiar with observational selection for quite some time, and some aspects of it have been being the subject of detailed mathematical modeling. However, it has never received so comprehensive and (in literal sense) universal a treatment as in this book.

The treatment is an excellent reply to those attacks (mainly ideologically motivated, coming either from the vulgar-materialist or positivist wing) on anthropic reasoning that dismiss it as anthropocentric, teleological or even quasi-religious.² At the same time, as already noted, defenders of the anthropic principle(s) are often motivated by this misinterpretation. One cannot help but wishing this ideological layer to be highlighted and banned from the serious discussion of these issues, a task Bostrom accomplishes with ease. In addition, he appreciates how the notion of the *multiverse* (or the *world-ensemble*), gaining ground in both cosmology and quantum mechanics, helps us understand anthropic “coincidences” as manifestations of selection effects. Here, the author is not only on the track of the distinguished contemporary analytic philosophers, like David Lewis or Robert Nozick, but also the key figures in modern quantum cosmology, notably Andrei Linde, Stephen Hawking, Alexander Vilenkin, Brandon Carter, or Don Page. Parenthetically, he dispells some misconceptions in the recent philosophical liter-

ature about the nature and validity of the explanations of *prima facie* improbable observations via multiverse. This would in itself be sufficient reason for writing (and reading!) of this book, but it is just the beginning.

The second part of the monograph (roughly Chapters 4 through 9) deals with statistical – in particular, Bayesian – approaches to the anthropic selection effects, and problems such approaches entail. Most of the funny and instructive thought experiments belong to this part of the book. It is the most “philosophical” part of discourse, invoking many elements of modern epistemology, probability theory, and logic. It is probably the most thorough and detailed application of the entire apparatus of Bayesian methodology to the case of our cosmological cognition. At the same time, Bostrom is unorthodox enough to surprise even the cognoscenti with a couple of new angles and unexpected twists. The most celebrated issue here is the (in)famous Doomsday argument (DA) which, in light of its possible consequences, deserves a slightly more detailed description. DA was conceived (but not published) by the astrophysicist Brandon Carter in the early 1980s, and was first exposed in print by John Leslie in 1989³ and in a *Nature* article by J. Richard Gott.⁴ The most comprehensive discussion of the issues involved is Leslie’s monograph of 1996, *The End of The World*. The core idea can be expressed through the following urn-ball experiment. Place two large urns in front of you, one of which you know contains ten balls, the other a million, but you do not know which is which. The balls in each urn are numbered 1, 2, 3, 4, . . . Now take one ball at random from the left urn; it shows the number 7. This clearly is a strong indication that the left urn contains only ten balls. If the odds originally were fifty-fifty (identically-looking urns), an application of Bayes’ theorem gives the posterior probability that the left urn is the one with only ten balls as $P_{\text{post}}(n = 10) = 0.99999$. Now consider the case where instead of two 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 the human race will soon become extinct (or at least that the number of individuals will be greatly reduced), and as a consequence the total number of humans that will ever have existed is about 100 billion. The other model indicates that humans will colonize other planets, spread

through the Galaxy, and continue to exist for many future millennia; we consequently 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 birth rank is about sixty billion. According to Carter and Leslie, we should reason in the same way as we did with the urn balls. That you should have a rank of sixty billion in the sequence of all humans is much more likely if only 100 billion humans will ever have lived than if the number was 10^{18} . Therefore, by Bayes' theorem, you should update your beliefs about mankind's prospects and realize that an impending doomsday – or at least a large decrease in the size of human population – is much more probable than you previously thought.

Bostrom investigates and rejects several objections to this, highly controversial, line of reasoning. He argues that DA is much stronger than people usually think when exposed to it for the first time, and that it does not fail for trivial reasons (an example of trivial objection: couldn't a Cro-Magnon man have used DA in his reasoning?). However, Bostrom does not accept the gloomy DA conclusion at face value, postponing (as any novelist worth his salt) the solution to the last part of the book.

Finally, the climax of the drama comes in the Chapter 10, which – together with the last chapter, the aftermath – expounds the new theory. As in any good detective story, the main culprit is finally revealed at the very end of the long thread of evidence: the universal observational selection effect, explained in detail in the Chapter 10. Here, Bostrom develops a theory which promises a unifying treatment of observations, in particular in cosmology, explicating in detail the accompanying Bayesian methodology. The central piece of it is the Observation Equation (p. 173), which subsumes seemingly vague assumptions and observational selection criteria in full mathematical rigour.

The unity of the underlying analysis is emphasized in the final Chapter, where the new theory is applied in several fields of contemporary research. From the Observation Equation it is possible to derive various anthropic results as special cases. Among several important contributions here, probably the most important one for physicists, cosmologists, and even astrobiologists is the solution of the “freak-observer” problem. Namely, in the absence of a compre-

hensive “Theory of Everything”, there are processes which are considered random, like the Hawking evaporation of black holes. In the infinite time of an ever-expanding universe (or *eternally inflating* multiverse!) these processes will unavoidably create some observers *without preceding evolution* which we are accustomed to link to observership from our own experience. For such “freak observers”, there will be no necessity of observing the delicately fine-tuned parameter values on which the anthropic coincidences are based. Would that invalidate statistical reasoning in cosmology? Bostrom, on the basis of the Observation Equation, answers this decisively in the negative, and his argument is robust and compelling. Similarly, he offers a solution for DA; but not to disclose too much, we leave to the reader to assess the strength of this newest reply to the big puzzle. And, of course, we get a new perspective on such diverse and disturbingly open issues as the thermodynamical arrow of time and the origin of traffic jams.

Probably the worst thing one can say about this book is that it is too short. After finishing it, the reader is left with the impression that the very scope of the new theory is such that there is enough material for an entire new book, or at least for a reconsideration of many issues treated in previous chapters. For instance, one cannot help feeling that the connection of the new theory to such important problems as the ongoing search for non-human intelligence (both SETI and AI projects) could be fleshed out in more detail. Nascent fields like astrobiology, artificial intelligence, or physical eschatology are excellent arenas for testing and applying the theory of observer-selection, yet we will not discern it from Bostrom’s dense discourse. In addition, the problem of the reference class (“who, or what, counts as an observer?”), remains highly controversial. At other places, Bostrom is quick to accept some rather fashionable – but nonetheless very uncertain – ideas, such as the strong AI thesis, without indicating that these are still remote from becoming parts of our established knowledge. But even these weaknesses actually serve more as a motivation for further thinking and work in this interdisciplinary field. In a sense, they increase the re-reading potential of this book. The reader will also find some consolation for finishing the book in a detailed and cleverly composed bibliography. All in all, *Anthropic Bias* is a wonderful achievement, which should

find place on the shelf of every serious student of modern philosophy of science, epistemology, and cosmology.

NOTES

1. I.e., J.D. Barrow and F.J. Tipler: 1986, *The Anthropic Cosmological Principle*. New York, Oxford University Press; Yu. V. Balashov: 1991, Resource Letter AP-1: The Anthropic Principle, *American Journal of Physics* 59: 1069–1076; J. Leslie (ed.): 1990, *Physical Cosmology and Philosophy*. London: Macmillan Publishing Company.
2. A recent example of such a view is: R. Klee: 2002, The Revenge of Pythagoras: How a Mathematical Sharp Practice Undermines the Contemporary Design Argument in Astrophysical Cosmology, *British Journal for the Philosophy of Science* 53: 331–354.
3. J. Leslie: 1989, Risking the World's End, *Bulletin of the Canadian Nuclear Society* 21 (May): 10–15.
4. J.R. Gott: 1993, Implications of the Copernican Principle for Our Future Prospects, *Nature* 363: 315–319.

Milan M. Ćirković
Astronomical Observatory Belgrade
Volgina 7, 11160 Belgrade, Serbia and Montenegro
E-mail: arioch@eunet.yu

