Milan M. Ćirković
Astronomical Observatory of Belgrade,
Future of Humanity Institute (Oxford University)
mcirkovic@aob.rs

Original Scientific Paper
UDC: 004.8:167/168
004.8:121

WHO'S REALLY AFRAID OF AI? ANTHROPOCENTRIC BIAS AND POSTBIOLOGICAL EVOLUTION

Abstract: *The advent of artificial intelligence (AI) systems has provoked a lot of discussions in both epistemological, bioethical and risk-analytic terms, much of it rather paranoid in nature. Unless one takes an extreme anthropocentric and chronocentric stance, this process can be safely regarded as part and parcel of the sciences of the origin. In this contribution, I would like to suggest that at least four different classes of arguments could be brought forth against the proposition that AI – either human-level or superintelligent – represents in itself a credible existential threat to humanity in either nearby or distant future. Part of the same argumentation is applicable to the general notion of postbiological evolution, which has caused even more unnecessary confusion and uproar in both laymen and philosophical circles. While the due attention should be given to the risks and dangers of the transitional period, there are many reasons why we should openly support and cheer for humanity's transition into the postbiological epoch.*

Keywords: *evolutionary theory, bioethics, artificial life, artificial intelligence, astrobiology, postbiological evolution*

> Whenever the word 'origin' is used, disbelieve everything you are told.
>
> Sir Fred Hoyle
>
> I think we agree, the past is over.
>
> George W. Bush

1. Introduction: the existential fear of AI

There are people who prefer not to see the link between artificial intelligence (AI) in its multiple forms and the sciences of the origin. This is unfortunate from both theoretical and practical points of view. First, and rather obvious, point is that "sciences of the origin" are not necessarily about the past. Unless one wishes to return to the pre-Darwinian or even pre-Copernican thinking, we should not regard ourselves and the world we are living in as pinnacles of creation and evolution. Quite to the contrary, we have little reason to think ourselves special in any way, including our position in the hierarchy of complex systems, i.e., in the abstract design space. To

 DOI 10.5937/BPA2235017C

some extent, this applies to our location in cosmic time as well. Although we have reasons to doubt the unthinking temporal Copernicanism (Ćirković and Balbi 2020), this still does not argue against futures origins events, especially when they concern close future and not the distant cosmological one. All in all, there should be nothing special about those kinds of events/processes we denote as "origins".

If we think about conventional origins as about the emergence/evolution of classes of complex systems, we are thinking about e.g.,

- (our) universe;
- (our) stellar system, i.e., the Milky Way;
- (our) planetary system, i.e., the Solar System;
- (our) planet Earth;
- (terrestrial) life;
- (human) mind;
- (human) language/culture.

Why stop there, however? The answer "because we don't see other classes of complex systems around us" is not only unphilosophical – it is also dangerously parochial and misleading. Not only should we contemplate things we are not in empirical touch with (otherwise, we would have never understood the atomic structure of matter, among other things), we should look more carefully around us. We have at least one process we actually are living through, which certainly qualifies as the origin: the emergence of various kinds of ultracomplex systems which go under the often too narrow label of artificial intelligence (henceforth AI). While the conservative understanding of AI suggests that, even when we achieve artificial *general* intelligence, it will be akin to human intelligence, this is not necessarily so – and might indeed be a dangerous conceit.

Further origins could as well lurk in our future – that is, located along the future temporal axis. We could also have origins which are removed from us in space, rather than in time. The formation of extrasolar planetary systems obviously belong to this category, as does the origin of extraterrestrial life, with the added spice in form of the panspermia hypothesis and other ways of introducing correlations between those origins. Those "other origins" deserve their own philosophical exploration, which is beyond the scope of the present paper. For the moment, we shall focus upon what is actually going on in the real world and what is putting the groundwork for the very next important "origins" – the key role of the emerging AI.

Clearly, human relation to AI is quite a complex and multifaceted topic. Detailed and proper analysis of it will constitute a large fraction of science and philosophy in the rest of this century, and likely in the centuries to follow. This aspect of the sciences of the origins will yet to be discussed and elaborated. Here, we focus on the very narrow aspect of the discussion which,

unfortunately, often stunts and impedes the really fruitful debate: correct construals of the concept of "AI risk" and our philosophical attitude toward it.

In the rest of this paper we shall first consider the physical eschatological argument for rejecting chronocentrism (Section 2), before we briefly survey the AI risk scene (Section 3) and outline classes of arguments to the effect that the AI risk is purely instrumental (Section 4). Some provocative tentative conclusions are given in the concluding section.

2. Physical eschatological argument

In order to properly assess our perspective on the very concept of "origins", we need to actively reject the bias of chronocentrism, defined as "the belief that one's own times are paramount" (Fowles 1974). There are multiple ways of doing so and my subjective preference for the arguments following from our global, cosmological knowledge should not occlude the fact that there are other, perhaps better, lines of attack.

One way of showing how outright ridiculous chronocentrism really is, leads through physical eschatology (Ćirković 2003). In brief, this young branch of physical science tells us that the future is way larger than the past. Some of the relevant future timescales are as follows (Adams & Laughlin 1997):

- Sun's future lifetime: 7.7×10^9 yrs;
- star-formation in the Galaxy continues for another $(5–10) \times 10^{10}$ yrs;
- longest-living stars existing today will exist for further 6.5×10^{12} yrs;
- the future lifetime of the Galaxy: cca. 10^{19} yrs.

Take just the last datum as a convenient placeholder. If it is correct – and there is no clear astrophysical reason at present why it should not be – *we are living in the first millionth of the first percent of our stellar system history!*

What's past is prologue. It's very brief, though. The fact that we are living so early in the course of the universal evolution – again, the first 0.000001% – of the lifetime of our stellar system, should give us a pause and perhaps instill some humility regarding our temporal position and epistemic pretensions. Plus, it should arguably motivate a reassessment of our research priorities, which has so far been extremely, staggeringly past-oriented rather than future-oriented.

Consider a man who has lived 80 years (not so rare these days, at least in developed countries) and all variables of his life. He might have been a sailor, a scientist, a criminal, a mailman, or whatever. He might have never married or might have been married three times with children. He might have been a passionate angler or biker or ecological activist or an atheist. Etc. etc. etc. Now, consider how much of this information on a rich and complex

life could you get by observing the first *5 minutes of his life*. Certainly not much. That is exactly the fraction of the total lifetime of the Galaxy we study as history this far. Therefore, it is exactly within this minuscule fraction that we have defined the origins.

Of course, the richness of the phenomena is not a linear function of the time elapsed. There are clear physical reasons to expect that the universal increase in entropy mandated by the Second Law of thermodynamics will make far-future epochs much slower as far as the evolution of various systems is concerned. Therefore, long future epochs of gradual decay of structures are much less interesting than the early epochs like the one we are living in. But even if we accept this effect of diminishing returns, interest-wise, the future is still so much larger than the past, even if we neglect plethora of very slow processes which have been unimportant so far but will become important in the future. We also need to take into account more bizarre and speculative cosmological futures we are currently vaguely aware of, such as the possibility of future vacuum phase transition at very low temperatures (e.g., Kusenko and Langacker 1997). In addition to all this, the actions of intelligent and intentional beings bring an entirely new dimension to our studies of the future, making some of our favorite inductive strategies largely irrelevant (Walker and Ćirković 2021). In fact, it is likely that only the development of AI will enable us to make sufficiently rich and detailed models of future evolution of complex systems including future human/posthuman societies; thus, the topic has both aspects of inevitability and self-reference.

3. AI fear-mongering: philosophy and marketing

If we allow for future origins, it is only natural to conclude that the first impending of such events is the emergence of AI and the accompanying onset of *postbiological evolution*. It is immaterial for the present argument which of the many construals of postbiological is the best or even acceptable; as a placeholder, we may refer to the convergence of nanotechnology, biotechnology, information technology and cognitive science (NBIC) to improve human performance (e.g., Canton 2004; Bainbridge and Roco 2006). So, while specifics may vary, we may expect that the NBIC convergence describes satisfactorily well human/terrestrial approach to the postbiological era. One important note must be made here: the locution "evolution" too often connotes blind Darwinian processes. While it may certainly be true that the classic Darwinian selection processes will continue to be in play among cybernetic posthumans or within the digital substrate, there are all reasons to expect the *mechanisms* of change to be dramatically different. If anything, postbiological evolution as a macroevolutionary trend could be expected to be Lamarckian, where intelligent agents consciously and deliberately choose their own evolutionary course, starting from an inherited state.

That would be the end of the discussion... if not for already more than a century of anthropocentric fear-mongering, starting with Karel Čapek's magnificent drama *R.U.R.* back in 1921. The spectre of "machine revolt" as the true *Phantom Menace* of the third millennium is certainly a global cultural phenomenon. One could trace its origins to myths and remythologized ideas such as Golems, Frankenstein monsters and in the modern context – robots. Moreover, the current phantasm implies not only that *Eeeevil Machines* are threatening us with existential disaster, so we ("obviously") have to "do something". If primitive bacteria in some pond 2 billion years BC had been able to think, they would have thought something along the lines *look at those pesky eukaryotes, they're threatening us, we Have To Do Something!* And some philosophers happily joined in the fear-mongering festivities, on a wide front from Robert Sparrow and Phil Torres to Toby Ord and Nick Bostrom.

This does not mean that there is no AI risk; the most comprehensive serious meta-survey up to 2014 is given by Sotala and Yampolskiy (2014). We need to be realistic about its magnitude and structure, however. There is no reason – at least not *a priori* – to expect the AI risk to be outside of the domain of the classical risk analysis, or to somehow circumvent the usual research approaches relevant for its assessment and management (cf. Scheessele 2021).

In that spirit, consider the distinction between substantive and instrumental risk in risk analysis. This is a distinction applicable to the anthropogenic/technogenic risks, such as the risks from nuclear weapons, synthetic biology or nanotechnology. Clearly, there are technologies which are carriers of *substantive risk* in and of themselves; nuclear weapons are almost perfect example in this regard. While some contrived peaceful applications of nuclear weaponry have historically been suggested (e.g., canal-digging), usually by people having vested interest in improving their public image, even that was not really assumed to be risk-free.[1] So, the concept of substantive risk applies to any kind of use. In contrast, instrumental risk is associated with the technologies which have legitimate risk-free uses. Instrumental risk goes under the vernacular terms such as „misuse", „abuse" and so on. Its inherent property is that it requires malevolent actors – and its key feature is that it removes nonsensical labeling of entire technologies as inherently risky or dangerous. One can obviously misuse all technologies, starting with the fire and the wheel, but it is on an entirely different level than the threat of nuclear weapons or synthetic pathogens.

1 Obviously, there is no risk-free activity *sensu stricto*. However, there is a clear and rather discrete difference between digging of a canal using shovels and digging of a canal using nuclear warheads, discernible even to philosophers.

4. Realism about AI risks and the arguments for it

At long last we come to the main thesis of this essay, which highlights the key feature of the incoming origins process: *the real AI risk is 100% instrumental risk*. Obviously, the qualification "real" is necessary here, since we need to assess what is truly at play here: clearly not what is sometimes hysterically *claimed* as the AI risk, but what can be realistically expected from the deployment of specific AI technologies. We should think about wild cards, of course: currently unimagined scenarios which increase the risk. We should not, however, allow for those wild cards to dominate the discourse.

There are several classes of arguments for the main thesis, which we shall outline in the following subsections.

4.1. Postbiological evolution

The most important argument comes from evolutionary biology and it is rather surprising that it has been very little discussed in the literature. Since the early days of the Modern Synthesis (in particular the work of J. B. S. Haldane which prefigured later results of George Williams and William Hamilton), if not from Darwin's *Descent of Man* in 1871, it was obvious that human behavioural traits are evolutionary grounded (Tattersall 2009). Therefore, as a matter of principle, the transition to postbiological phase obviates most, if not all, biological motivations. The very definition of ecology and the relevant ecological needs and imperatives changes, leading to significant changes in other fields which have been traditionally linked to the evolutionary processes.

As an example, the imperative for filling the complete ecological niche in order to maximize one's survival chances and decrease the amount of biotic competition is an essentially biological part of motivation for any species, including present-day humans. It would be hard to deny that this circumstance has played a significant role in colonization of the surface of the Earth. But expanding and filling ecological niches are not intrinsic properties of life or intelligence – they are just consequences of the predominant evolutionary mechanism, i.e., natural selection (e.g., Tinbergen 1968; Wilson 1978; Trivers 1985; Heying and Weinstein 2021). It seems logically possible to imagine a situation in which some other mechanism of evolutionary change, like the Lamarckian inheritance or genetic drift, could dominate and prompt different types of behaviour. The same applies for the desire to procreate, leave many children and enable more competitive transmission of one's genes to future generation is linked with the very basics of the Darwinian evolution. Postbiological civilization is quite unlikely to retain anything like the genetic lottery when the creation of new generations is concerned. In addition, the easiness of producing and retaining copies of postbiological organisms in the digital substrate is likely to dramatically change the meaning of terms such as

"maturation", "adulthood", "parenthood", "kin", etc. Thus, we need to make an additional step symbolically represented as the analogy:

biological evolution → postbiological evolution
sociobiology → "post-sociobiology"
biologically rooted behaviors → behaviors based on postbiological factors.

Clearly, we need much more research and thinking in order to establish what exactly could "post-sociobiology" be, but as a provocation we may suppose that it will deal with "stable ingredients" (to use the expression of Arnold Toynbee; see Toynbee 1966) of postbiological development. In the case of (post)human evolution, one may argue that this will encompass "posthuman nature" in the same manner as authors like Fukuyama (2002) invoke "human nature" as an explanatory device. It is very hard to conceptualize such a dramatic change – but we still ought to think as hard as possible about its outcomes since, among other things, some very early decisions can have long-reaching consequences (Bostrom 2003b).

In brief, the biological evolutionary baggage such as aggression, teritoriality, tribalism, possessiveness, etc. etc., probably responsible for 99% of problematic features of human culture and history will be getting *weaker*, not stronger, with the advent of AI. No biological evolution whatsoever means no baggage! And, clearly, a postbiological-based ethics, whatever that is, should substitute for the current, biological-based ethics.[2]

4.2. The design space non-ergodicity

It is very reasonable to assume that postbiological design space is much, much bigger than the biological morphospace; after all, a part of the motivation for transition to the postbiological realm is exactly to mitigate the weaknesses and insufficiencies of the biological. If a cyborg with transplanted human brain or an uploaded mind can survive and thrive on the lunar surface, for example, this means that a new wide subspace of the postbiological design space will have opened, with no analogs in the space of biological forms. (Assumed, of course, that there are no biological forms which could survive on the lunar surface, which – while very plausible – cannot yet be regarded as definitely proven.)

This difference in generalized space size cannot come for free; the no free lunch axiom applies here as well. An important consequence of the difference in size is that trajectories are much more complex to navigate within larger

2 For those worrying excessively about the extinction of humanity as a consequence of "new and better" model of intelligent beings coming online, we should note that even within *purely biological* evolution, "nature red in tooth and claw", the emergence of new taxa (even high ones!) did *not* mean the removal of the old ones. Thus, we still have the sponges (*Porifera*), the comb jellies (*Ctenophora*), and other ancient phyla and other taxa.

than within smaller space. If one wishes to arrive at a desired point – or a desired region of the parameter space – one has to steer much more carefully in a larger space than in a smaller one. That much is a clear mathematical truth, not something abstract, hypothetical, or conjectural.

What is somewhat conjectural, but nevertheless highly plausible, is that taking evolution in our own hands is going to be extremely complex affair, which can hardly be calculated without a new, powerful numerical models. And this applies to any and all aspects of the transition, including the ecological ones. Postbiological minds are simply likelier to navigate the complex sustainability trajectories. (Which, parenthetically, apply to the extraterrestrial civilizations and the crossroads between astrobiology and the futures studies; cf. Frank and Sullivan 2014.)

4.3. "Wisdom of the ages"

It is often claimed, especially among conservative thinkers, that while *intelligence* of AI systems – or indeed many fellow humans – is measurable and indubitable, their *wisdom* is at the very best limited (cf. Fukuyama 2002). Now, this argument would have more merit if it were accompanied by an operational definition of wisdom. Unfortunately, this is usually not the case, so one has to improvise irrespective of whether one accepts the argument or not.

Suppose that we construe wisdom as a set of beneficial insights and statements tested by history and ages; beneficial, presumably, for both individual and the community she lives in. Clearly, in the human experience so far, it is very, very difficult to develop new or expand the existing wisdom, since the human history and the history of each culture and community, as well as life history of any individual are unique and do not allow for controlled experiments of the kind we have in most physical sciences. That very circumstance, that *new wisdom is hard to come by*, can be regarded as a major (or even *the* major) tenet of conservatism and perhaps the reason why conservatism is an anthropological constant over all ages and all cultures.

AI brings an entirely novel element in this perennial story. History is not perfectly immutable set of records any more. One can reasonably experiment with history via detailed simulations, potentially with arbitrarily wide scope and high resolution. The simulations of the kind envisioned here are extremely complex and similar to the concept of the "ancestor simulations" of Bostrom (2003a). There are all kinds of auxiliary problems and issues related to the existence of such simulations, but they are immaterial for the present goal. What is crucial is that such historical simulations enable experimenting with various evolutionary trajectories – and hence *maximizing wisdom*!

Not only that: the process of optimization could, in principle, be sped up. Postbiological "effective age" is a function of the processor clock rate. Therefore, experimenting with history will enable much *quicker* acquiring of wisdom (as operationally defined above). Multiple internal simulations

will likely bring insight into all kinds of moral dilemmas – which is the key ingredient in any kind of construal of wisdom. All in all, it could be persuasively argued that it is in fact likelier that future advanced AIs will be endowed by true wisdom than what we could say, for example, for our present political and cultural leaders.

4.4. Semantic dysfunction of "nature" vs. "artifice"

Ever since Turing proposed "a heretical theory" (Turing 1970) the specter of the "artificial Other" has been haunting human thinking. Insofar we accept physicalism about minds, the distinction between "natural" and "artificial" minds becomes blurry at best. There is no "natural" trade mark which could be stamped on a mental state if it is equivalent to a physical state of matter, governed by the universal laws of physics. There is no difference between properties of the water molecule produced in a laboratory by joining two hydrogen atoms with one oxygen atom and the H_2O molecule taken from, say, a river. Historically contingent origin of even those hydrogen and oxygen atoms themselves (hydrogen being produced in the Big Bang nucleosynthesis, oxygen mostly in low– and intermediate-mass Population II stars) does not impact their properties here and now. This applies, rather obviously, to their *yet undiscovered* functional properties, for instance, their participation in a not-yet-synthesized complex pharmaceutical compound. In more general terms, physicalism at least vaguely implies insensitivity to historically contingent initial conditions (even without taking into account specific physical *processes* like nonlinear dynamical evolution, which erase information on the initial states). *Why would, then, anyone assume the difference between properties of a mind produced in an AI lab and properties of a mind evolved by biological evolution?*

Whoever accepts physicalism and is still afraid of AI *in the substantive sense* is a bit of a hypocrite. Worrying that something especially bad or corrupt will happen because AI is labeled as artificial is dangerously similar to anthropocentric justification of crimes and corruption perpetrated by those minds labeled as natural.[3] One could speculate that the erroneous demarcation between the „natural" and the „artificial" is an evolutionary consequence in itself, as are most of other persistent irrational beliefs (e.g., religious ones). This should not discourage us, however, to problematize and criticize it at any point. In fact, this kind of argument is the true point of both Čapek's *R.U.R.* and Ridley Scott's celebrated movie *Blade Runner*: discrimination against robots or replicants on the basis of their allegedly "unnatural" origins is a dangerous anthropo-chauvinistic nonsense, which cannot end well.[4]

3 Obviously, non-physicalism about minds open a host of other possibilities, including those in which non-physical ingredient prevent artificial minds from being "true" minds, which is perhaps the closest to the folk understanding of AI in most of the world.

4 Notice the absent argument in all of the above, which has been sometime heard in transhumanist circles: that the AI risk is not really a risk, since even the adverse outcome will have net positive value. (This has been circulated much informally, but rarely in

5. Conclusions

We have so far, in spite of much hand-waiving, obtained not a single strong reason to believe that the AI risk is anything but instrumental. This especially pertains to the alarmist positions which argues that the instrumental fraction of the AI risk tends to zero or some very small quantity. Those positions betray self-serving, antievolutionary and anthropocentric attitude of their proponents and tell us little about the reality (or otherwise) of true AI, be it of human or superhuman level.

Philosophical failure to put the relevant concepts into the truly Copernican, non-anthropocentric, non-chronocentric evolutionary context should worry us enormously. It shows how empty our oft-verbalized proclamation of inclusiveness, universality, and brotherhood of the mind are in reality. In the times when astrobiologists prepare to search in detail for biosignatures on extrasolar planets, and the new generation of search for technosignatures (previously known as SETI), it is indeed deplorable that we are seemingly not able to adequately conceptualize other minds, *even if they are of our own making*. What to expect, then, of future contact with the independently-evolved extraterrestrial life and intelligence? If fear and paranoia dominate our thinking about minds we have built ourselves, what hope is there that we will be able to empathize with complex beings we have no phylogenetic relation to whatsoever? Even more, won't such fear and paranoia lead us to totalitarian oppression of our own kind to preserve "essential human dignity", "human essence", "neural purity" (cf. Reynolds 2006), and such demagogic, ideology-laden nonsense?

All what has been said should not be construed as stating that the transition to postbiological will be a cozy ride. On the contrary, there are all kinds of indications that – similar to the previous global revolutions such as the agricultural and the industrial revolution – it will be accompanied by much upheaval, societal chaos, and risk in general (Bostrom 2014; Ćirković 2017; Ord 2020). It is paramount, however, to understand where does that risk mostly come from: anthropocentric human institutions, from political and judicial systems, via economic structures, all the way to media, institutions of culture, etc. For instance, it has already become obvious that automatization and robotization are bringing about a surge in unemployment, as well as gradual obsolescence of many cherished human professions and trades. As we

---

print; perhaps the closest are formulations of Moravec 1988, 1998.) This is nonsense not only because the definition of risk can never be given entirely objectively which is a foundational principle of risk analysis since its emergence as a discipline (e.g., Byrd and Cothern 2000), and always refers to a set of subjects. It also manifests a pathology which could be justly dubbed the "utilitarian Stockholm syndrome": incorporating parts of the ethical mindset and values of one's captors on the basis of allegedly correct utilitarian calculus. Detailing fallacies inherent in this attitude would require a separate paper and is not of great interest for us here.

have seen in recent years, this has already caused social unrest in many of the affluent societies on the planet; we need to be wary and extremely cautious regarding further stages of this process, to avoid sliding into totalitarianism. The advent of the true general AI is likely to bring about now unconceived problems, true "unknown unknowns" we need to be agile in perceiving and analyzing.

All those, however, stem from the instrumental use of AI and related technologies and do not present us with something necessarily dangerous in and of itself. While putting AI in hands of dictators, supreme religious leaders, unscrupulous spymasters, mafia bosses, white-collar criminals, or similar unsavory characters is certainly a very, very bad idea, the badness is just a function of negative value assessment of those characters. Putting *anything* of value in hands of dictators et al., is a very bad idea as well; this is valid irrespectively of the specific nature of AI. While AI can magnify the evil-doing of evildoers by a large factor, the root cause is clearly the existence of evildoers among humans. While all this is embarrassingly trivial, the fact that many authors and even more people among the general public find it easy to slip into the "AI is evildoer itself" mode of *2001: A Space Odyssey* (1968) is quite disturbing.

For the very end, whoever finds scientific books and papers boring should read some *belles-lettres* about the postbiological evolution. Almost 30 years ago, the great Australian master of hard science fiction, Greg Egan, has published a majestic novel, *Permutation City*, which does not only deal with the postbiological evolution, but whose protagonists – at least in the second part of the book – are indeed cellular automata (Egan 1994)! Egan's novel not only highlights the capacity of such a transformative technology, but brings a dire warning of problems we may have with our own intellectual progeny – if we continue along the anthropocentric road. It is a common knowledge how difficult successful science and technology public outreach is. Therefore, one should use and celebrate each and any lump of gold encountered along the way.

Acknowledgements. The author is happy to express gratitude to the organizers of the *Sciences of the Origin: The Challenges of Selection Effects and Biases* conference (June 3–5, 2021), to an anonymous reviewer for useful comments, as well as to Vojin Rakić, Barry Dainton, Mark Walker, Srdja Janković, and Ana Katić for helpful discussions.

References

2001: A Space Odyssey [1968] 1997. DVD, dir. S. Kubrick, Metro-Goldwyn-Mayer, Beverly Hills, California.

Adams, F. C. and Laughlin, G. 1997, "A dying universe: the long-term fate and evolution of astrophysical objects," *Rev. Mod. Phys.* **69**, 337–372.

Bainbridge, W. S. and Roco, M. C. 2006, *Managing nano-bio-info-cogno innovations* (Springer, Dordrecht).

Blade Runner [1982] 2007. DVD, dir. R. Scott, Warner Bros., Burbank, California.

Bostrom, N. 2003a, "Are You Living in a Computer Simulation?" *Philosophical Quarterly* **53**, 243–55.

Bostrom, N. 2003b, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas* 15, 308–314.

Bostrom, N. 2014, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, Oxford).

Bostrom, N. and Ćirković, M. M. (eds.) 2008, *Global Catastrophic Risks* (Oxford University Press, Oxford).

Byrd III, D. M. and Cothern, R. C. 2000, *Introduction to risk analysis: a systematic approach to science-based decision making* (Rowman & Littlefield, Lanham).

Canton, J. 2004, "Designing the future: NBIC technologies and human performance enhancement," *Annals of the New York Academy of Sciences* **1013**, 186–198.

Ćirković, M. M. 2003, "Resource Letter PEs-1: Physical Eschatology," *Am. J. Phys.* **71**, 122–133.

Ćirković, M. M. 2017, "Enhancing a person, enhancing a civilization: a research program at the intersection of bioethics, future studies, and astrobiology," *Cambridge Quarterly of Healthcare Ethics* **26**, 459–468.

Ćirković, M. M. and Balbi, A. 2020, "Copernicanism and the Typicality in Time," *International Journal of Astrobiology* **19**, 101–109.

Dick, S. J. 2003, "Cultural Evolution, the Postbiological Universe and SETI," *Int. J. Astrobiology* **2**, 65–74.

Egan, G. 1994, *Permutation City* (Orion/Millennium, London).

Fowles, J. 1974, "On chronocentrism," *Futures* **6**, 65–68.

Frank, A. and Sullivan, W., 2014. Sustainability and the astrobiological perspective: framing human futures in a planetary context. *Anthropocene*, *5*, pp.32–41.

Fukuyama, F. 2002, *Our Posthuman Future: Consequences of the Biotechnology Revolution* (Farrar, Straus & Giroux, New York).

Heying, H. and Weinstein, B. 2021, *A Hunter-Gatherer's Guide to the 21st Century: Evolution and the Challenges of Modern Life* (Portfolio, London).

Janković, S., Katić, A., and Ćirković, M. M. 2022, "Gaia as Solaris: An Alternative Default Evolutionary Trajectory," *Origins of Life and Evolution of Biospheres*, in press (DOI: 10.1007/s11084–022–09619–6).

Kusenko, A. and Langacker, P. 1997, "Is the vacuum stable?" *Physics Letters B* **391**, 29–33.

Moravec, H., 1988. *Mind children: The future of robot and human intelligence*. Harvard University Press.

Moravec, H. 1998, *Robot: Mere Machine to Transcendent Mind* (Oxford University Press, Oxford).

Ord, T. 2020, *The precipice: Existential risk and the future of humanity* (Hachette Books, New York).

Reynolds, A. 2006, *Galactic North* (Gollancz, London).

Scheessele, M. R. 2021, "The hard limit on human nonanthropocentrism," *AI & SOCIETY* **37**, 49–65.

Sotala, K. and Yampolskiy, R. V. 2014, "Responses to catastrophic AGI risk: a survey," *Physica Scripta* **90**, 018001.

Tattersall, I. 2009, "Charles Darwin and human evolution," *Evolution: Education and Outreach* **2**, 28–34.

Tinbergen, N. 1968, "On war and peace in animals and man," *Science* **160**, 1411–1418.

Toynbee, A. 1966, *Change and Habit* (Oxford University Press, Oxford).

Trivers, R. L. 1985, *The Social Evolution* (Benjamin Cummings, Menlo Park).

Turing, A. 1970, "Intelligent Machinery, A Heretical Theory: Report, National Physical Laboratory," *Machine Intelligence* **5**, 3–23.

Walker, M. and Ćirković, M. M. 2021, "Induction Ain't What It Used to Be: Skepticism About the Future of Induction," *Journal of Ethics and Emerging Technologies* **30**(1).

Wilson, E. O. 1978, *On Human Nature* (Harvard University Press, Cambridge).