BOOK REVIEW

Here be Dragons: Science, Technology and the Future of Humanity, by Olle Häggström, Oxford University Press, 2016, 278 pp., $44.95 (hardcover), ISBN 978-0-19-872354-7

Citation for this article:
Haqq-Misra (2016) *Law, Innovation and Technology*, doi:10.1080/17579961.2016.1250381

Scientific and technological progress holds the potential to radically transform the future development of civilization. Advances in artificial intelligence, atomically precise manufacturing, human enhancement, and climate engineering could offer breakthrough solutions in the near-term that can help to solve major global challenges. Yet the capabilities implied by these technologies also invoke new sources of risk that could cripple or destroy the foundations of civilization. The juncture of contemporary research at the frontiers of knowledge and speculations about future risks forms the foundation of Olle Häggström's book *Here be Dragons*, which offers a broad survey of major topics in future studies and global catastrophic risk. Häggström offers his perspective that future developments in science and technology could potentially make us much worse off, suggesting that new modes of thinking about the direction of future progress are needed to help ward off potential catastrophe.

The book begins with an overview of problems at the intersection of science and policy, arguing that some problems in contemporary science cannot be cleanly divorced from their ethical counterparts. The discussion then diverges into an overview of the Earth system and its major variable factors, which places into context the current climate change problem. All of the most prominent contemporary climate engineering proposals are given consideration, and Häggström further raises the possibility of 'moral bioenhancement' to aid in solving the commons problem of anthropogenic climate change. This leads into a general discussion on prospects for human enhancement through genetic engineering, brain-machine interfaces, and other possibilities that remain consistent with the trajectory of contemporary research. The general history of computing that follows allows for a full appreciation of the possibility for artificial general intelligence and the risk of an intelligence explosion, followed by a foray into the benefits and risks of atomically precise manufacturing. The focus of the book shifts midway through toward philosophical issues, beginning with an overview of major questions in the philosophy of science. The doomsday argument as well as other arguments for global catastrophe are considered in turn, concluding with a discussion of space colonization, intergenerational discounting of costs and benefits, and possible directions for future research.

Perhaps surprisingly, a major strength of the book comes from the author's background as a professor of mathematical statistics. Of course, Häggström provides the necessary concise analyses of some of the major philosophical issues that underlie efforts such as human engineering, artificial intelligence, and nanotechnology, and he offers precise distillations of broad literature into a collection of the most important risk elements. Yet his probabilistic insight throughout the book is reflected in his prose as he remains agnostic about whether or not

possibilities such as whole-brain emulation, cryonics, or atomically precise manufacturing will ever be possible; yet he is still willing to subjectively rank the associated risks in these emerging fields for the purpose of prioritizing preventative measures. Häggström explicitly discusses the effective use of Bayesian statistics for learning about probabilities midway through the book, relegating most of the detailed mathematics into footnotes, but allowing the reader to grasp some of the subtleties and possible inconsistencies that arise with Bayesian reasoning. The use of examples and probabilities in the main text is met with articulate explanations, and so even an undergraduate student armed with algebra should be able to access Häggström's arguments.

Häggström also provides a much-needed exposition on the context of falsification in the philosophy of science. Karl Popper's criterion of falsification is often dismissed by practitioners of Bayesian statistics, but Häggström demonstrates that both methods offer tools that are still relevant today. This includes a discussion of the discovery of Neptune, which is often cited as a counterexample to Popperian falsification; yet Häggström correctly identifies that such a naive interpretation need not be the only approach toward falsification. More generally, Häggström elaborates upon the distinction between frequentist and subjective probabilities, offering a similar interpretation that both should be regarded as tools to match the problem at hand. This balanced view of these perspectives in probability theory and the philosophy of science is among the greatest assets of the book.

As the book descends into more speculative territory, Häggström's personal opinions begin to show, perhaps to the loss of some objectivity. He raises the Doomsday argument put forth by astrophysicist Brandon Carter and others, and he offers a complete mathematical treatment of the problem from both frequentist and Bayesian perspectives. This construction of the problem is sound, but Häggström then declares that the Doomsday argument is nothing more than a distraction and should be ignored. The reader is left with the uncertainty of a well-defined problem that at least holds relevance to problems in anthropic reasoning, if not actual doomsday, and the arguments for dismissal are given no more force than the conviction of the author's opinion. However, the footnotes contain sufficient references to the relevant literature that any interested reader could easily continue their study of the Doomsday argument. Understandably, Häggström's purpose is to focus on the reduction of future risks, so in this sense the Doomsday argument may indeed be somewhat irrelevant to the implementation of actual risk reduction measures.

The concluding sections focus on possibilities for space colonization and the search for extra-terrestrial intelligence (SETI) elsewhere. He discusses both the conventional search for radio signals transmitted by extra-terrestrials as well as prospects for discovering evidence of macro-scale engineering in the galaxy; Häggström opines that he favours the latter will be more successful, although such an assertion is difficult to justify. He then raises the question of whether or not humanity should 'shout into the cosmos' by engaging in messaging to extra-terrestrial intelligence (METI) as a counterpart to SETI. He calls such activities 'inexcusably reckless' and cautions against any such activities now or in the near future. While such an opinion can be justified given our lack of knowledge of extra-terrestrial life in the galaxy,

Häggström's discussion of METI falls short of the objectivity he offers to other topics. Most notably is that Häggström raises several key questions that are critical of METI, yet his exploration of possible responses to these questions that might result in a neutral or pro-METI stance overlooks arguments that have been discussed since the SETI program first began. The literature review and footnotes in this section also indicate a selection toward the author's apparently preferred conclusion and offer a less than complete survey of contemporary thinking on the topic than other sections of the book. Häggström's treatment of METI could have been improved by analysing the relative risks of engaging in METI versus not engaging in METI; such an analysis might still conclude that we should not engage in METI today, but it would provide a more transparent basis for preferring such a conclusion.

The conclusion of the book was admittedly disappointing and failed to provide a concrete direction for how to address the major problems outlined in the chapters earlier. Häggström offers a discussion of economic discounting and its application toward thinking about the future, and he then relates this to our prioritization of existential threats in the distant future. This adequately captures the magnitude of the problem between our current modes of short-term thinking and the need for longer-term risk reduction, but Häggström offers only vague ideas about how to proceed. The best advice he can give is for us to 'map those territories' that are uncertain, where the metaphorical dragons on our atlas reside. This may indeed be the best one can do from such a wide survey of topics, and Häggström also provides a list of critical questions that can help guide thinking about the relative value of present and future people and environments. Yet Häggström's concluding thoughts lack the type of systematic prioritization of risks and reduction methods that could have arisen out of such an analysis.

Here be Dragons provides an overview of topics at the intersection of future studies and global catastrophic risk. The rich bibliography provides a valuable reference for any student or professional in the field. Although the book also could serve as an entry point for anyone new to the field, the somewhat technical discussions and occasional liberal use of opinion may make this better suited as a secondary, rather than primary, source. The book is suitable for an undergraduate or graduate level audience as an introduction to the major technological and philosophical issues that are emerging today.

Jacob Haqq-Misra
Blue Marble Space Institute of Science
jacob@bmsis.org