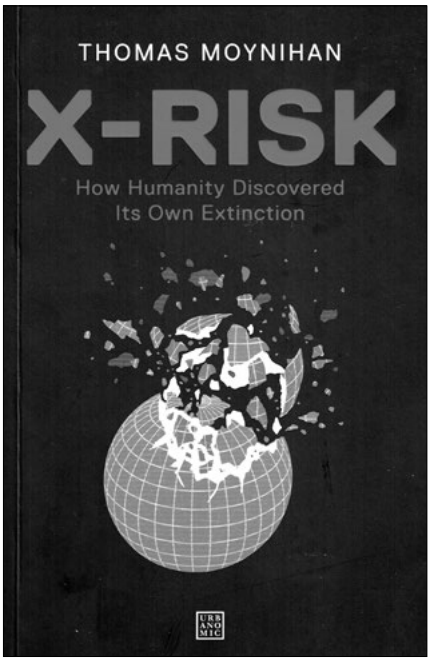
Thomas Moynihan: X-Risk: How Humanity Discovered its Own Extinction

Reviewed by Kritika Maheshwari



Technology experts are now claiming that superintelligent artificial intelligence, if realised, could pose an existential threat to humanity's long-term survival. The possibility that we might be putting humanity at risk of extinction has instantly spurred world leaders into action, with some insisting we put a pause on research and training of artificial systems. These recent events might perhaps suggest that humanity is finally waking up to the realisation that our future is anything but secure. However, as Thomas Moynihan argues in his recent book *X-Risk* (2020), this concern with existential threats has a long intellectual history, one that is important for understanding how and why we ought to care about humanity's continued survival into the future. Writing as a historian and philosopher of ideas, Moynihan carefully weaves together a complex account of how humanity first discovered the idea of its own extinction. By drawing on historical, literary, philosophical, and theological perspectives on the topic, Moynihan makes three claims along the way.

First, Moynihan makes the case for the novelty of humanity's extinction. He argues that the idea of human extinction has remained conceptually unavailable for the most part of our existence as a species. Next, he argues that the apocalyptic prophecies one reads about in religious and mythological texts are both conceptually and normatively distinct from the idea of human extinction. Whereas the thought of apocalypse offers a sense of an ending and is thus a conciliatory concept, the idea of our extinction anticipates the ending of sense and rationality and thus offers no consolation. Finally, Moynihan contends that fully grasping the prospect of our own extinction is not to be celebrated merely as a conceptual feat. Instead, we must recognise our ability to reason about humanity's extinction as a defining feature of modernity itself. Drawing upon philosophical ideas marshalled by Enlightenment thinkers like Immanuel Kant, Moynihan argues that our own rationality draws attention to the responsibility we have to ensure humanity never meets the disastrous fate of going out of existence.

Understanding how exactly the Enlightenment period succeeded in placing existential risk on humanity's conceptual map first requires a brief historical detour into ancient thought. In chapter 2 (*Cosmic Silences: Astrobiology)*, Moynihan emphasises the stronghold of a pre-Enlightenment philosophical assumption, namely the principle of plenitude which states that all legitimate possibilities in the world are realised. The idea of plenitude entails that should our species go extinct, the possibility of its return will eventually and inevitably be fulfilled. This plenitude-centred thinking dating back to ancient philosophers like Plato, Pliny, and Lucretius has the following upshot: It suggests that moral understanding and moral justification for humanity's extinction relies heavily on what we accept as the correct or appropriate metaphysical and scientific view of the world, all other things being equal. So, if it's true that humanity would reappear once extinct as a matter of necessity, then the question of whether causing or allowing humanity's extinction is morally wrong loses its significance.

Historically, then, the prominence of plenitude-centred thinking, together with the now frequently rejected suggestion that nature itself is imbued with value and justice, dismissed the case for even thinking about human extinction by rendering the very idea of extinction meaningless. This brief yet important insight into pre-Enlightenment thinking about our future, or rather the absence of it, is interesting. It stands in sharp contrast to our recent preoccupation with mitigating and strategising about different existential risks that face humanity today. This indicates that we have moved intellectually from an adherence to plenitude to an acknowledgement of the contingency of the conditions of human existence and the role of chance, which raises the question of how humanity came to acknowledge extinction as an issue worthy of its attention?

In chapters 3 (*Earth Systems: Geoscience*) and 4 (*Future Trajectories: Forecasting*), Moynihan examines how distinct fields of empirical science such as geosciences and actuarial sciences converged upon the idea that we hold the power to either push humanity to the brink of a precipice or use that power to preserve our long-term future. The intellectual shift towards this Enlightenment frame of thinking was marked by rejecting the otherwise widespread conflation of moral values with natural facts. This entailed a further radical shift in our thinking about our collective future, namely that it is not only open and uncertain, but also precarious. In chapter 3, Moynihan reviews how scientific studies of fossils as well as species mutability provided empirical evidence for the reality of species extinction. The Leibnizian idea that ours is the best possible world was soon questioned by the reality of pre-historic non-human extinctions, opening up doors to the possibility that humanity's continued existence is a mere accident. Moreover, evolutionary principles such as Louis Dollo's law of irreversibility reified the idea that even if plenitude is plausible, humanity's extinction would be irreversible insofar as organisms can never return to their former state even when placed in identical conditions to those in which they previously thrived.

In *Future Trajectories: Forecasting*, Moynihan offers another good example of how advancement in scientific theorising has strongly shaped our present understanding of extinction, both as a natural and a moral phenomenon. He turns his attention towards political arithmetic, or rather demographic thinking, which sowed the seeds for considering humanity as a object for objective investigation. Thinking about humanity as an aggregate may not initially strike us as an impressive feat because we are now so used to population-level thinking in matters of policy and political decision making. However, this was an achievement par excellence during the Enlightenment period, for it allowed us to conceive of humanity as a planetary collective. Combined with progressions in our mathematical understanding of risk, probability, and uncertainty, it was now also possible to have a quantitative grasp of existential threats humanity at a collective scale.

The Enlightenment period was thus successful in reinforcing the idea that our extinction would lead to the end of all value. The decoupling of fact from nature, the dismissal of plenitude, as well as empirical evidence for existential risk suggested two potential approaches to this dilemma: either we take nature's lack of inherent prudence and morality as an engineering problem in need of a fix or we dismiss any responsibility we may have towards preserving and protecting our collective future.

In chapter 5 (*Internal Contradictions: Omnicide*), Moynihan explores a range of views advocating for latter approach on three key bases: that perhaps our concern with human existence justifies a problematic kind of human exceptionalism, that perhaps living in the worst possible world full of suffering and sufferers is a live possibility, and that perhaps extinction is in fact the key to unleashing rather than curtailing our potential. Are the moral stakes involved in our extinction great enough to outweigh the harms of human exceptionalism, suffering, and curtailing our own potential?

Moynihan's project is not to settle these issues, but rather is to explain how we came to care about humanity's precariousness in the first place and to suggest why we must continue to care. In chapter 6 (*Physical Salvation: Vocation*), Moynihan argues that the answer is to be found in Kant's philosophy and in particular, in the idea that moral values are a question of self-legislation. In arguing against the idea that values are inherently imbued in nature, Kant argues that they are maxims that we elect to bind ourselves to and are thus our own responsibility. Hence, part of being a rational actor is to become concerned by the extinction of rationality, for it cannot exist otherwise. It is in this sense that as rational actors, we were bound to discover humanity's extinction through our ability to act and think rationally. Equally, we are responsible for caring and doing something about the existential risks we face. As such, dismissing existential risk on the basis of plenitude or on accounts of conflating fact with nature is simply incoherent with the bounds of Kantian thought. Moynihan's recasting of the origins and importance of existential thinking from this perspective is original and an important contribution to the project of developing a Kantian ethics of human extinction within a theoretical landscape that currently remains dominated by consequentialist theorisation. In doing so, Moynihan takes the first step towards providing an *explanation* for why it is rational for humanity to care, or rather, continue to care about its own extinction. However, it a separate question whether this explanation also provides us with a comprehensive justification of humanity's attempt to prevent its own demise.

For instance, let us suppose that humanity's future can be only protected by willing the end of all non-rational life on Earth. Within the constraints of an anthropocentric theory which is committed to the idea that rational nature alone has absolute and unconditional value, mitigating existential risks this way may not seem dismal. And yet, many would abhor the idea of preserving our rationality at the cost of sacrificing or destroying everything else we may value, such as beautiful landscapes, trees, and non-human animals. Similarly, what if avoiding humanity's complete self-annihilation required the self-annihilation or moral suicide on part of some rational agents? In which ways can Kantian ideas of perfect duties to the self as it applies both to individuals and humanity as a whole guide us? Such examples are merely intended to show that observing the moral significance of caring about humanity's extinction through Kantian lens raises new questions about *how* we ought to care. Moreover, it also raises questions for *whom* we ought to care for.

For example, let us consider the project of reconciling Kantian ethics of extinction with the prominent consequentialist thought that causing or risking our extinction is morally wrong as it blocks the added value of bringing future people into existence. As Moynihan notes, "to give up the fight to maximise value is to immorally submit to the environing forces of extinction, to the unjust fact that extinction and sterility is the cosmic tendency and the uphill struggle toward complexity the exception" (367). This, however, raises the question of whether and in what ways the Kantian injunction to respect the autonomy of actual persons rules out or alternatively includes potential people within the scope of its moral community, to whom we owe this concern. Again, the point here is not to dismiss Moynihan's claim that humanity's concern for its extinction is presupposed by the very nature of rational agency itself. Rather, it is to motivate further investigations into how far we can take this idea and apply them to concerns that occupy those interested in ethics of our long-term survival.

As Moynihan correctly notes, this Enlightenment-driven idea is still a work in progress – we are only now starting to uncover the full ramifications of humanity as historic collective project. This process remains incomplete both because we are far from achieving humanity's full potential, but also with regards to reifying the scope and the content of responsibility that rationality places on individuals for mitigating the existential risks that humanity faces. A few important questions should be raised in this context: What is our individual responsibility towards mitigating such risks? How does our individual responsibility fare against our collective responsibility as a rational species? Besides, what demands are placed by rationality onto the preservation of rationality itself? For instance, would humanity's long-term potential be preserved if human life were to be replaced not by superintelligent, but some kind of superrational artificial intelligence? In conclusion, Moynihan's book not only succeeds in capturing the historical landscape of humanity's extinction, it also manages to push the boundaries of philosophical inquiry by raising new and important questions worthy of further research.

Moynihan, Thomas (2020): X-Risk. How Humanity Discovered its own Extinction. London: Urbanomic Media. 472 Pages. ISBN: 9781913029845. Price €25 (paperback).