Futures 116 (2020) 102495

Contents lists available at ScienceDirect

Futures

journal homepage: www.elsevier.com/locate/futures

Essays

Existential risk and human extinction: An intellectual history

Thomas Moynihan*

University of Oxford, United Kingdom

ARTICLE INFO

Keywords:
Existential risk
Intellectual history
Human extinction
Enlightenment
Geoscience
Probabilism

ABSTRACT

Of late, existential risks have become the target of an emerging field of scientifically serious study. This baptism of 'X-risk studies' is symptomatic of what Riel Miller has diagnosed as an ever-increasing demand for 'futures literacy', inasmuch as we are progressively conversant with progressively distal perils. Yet this dynamic, of incremental 'future orientation', is not itself without a history. We have been being swept up in the future *for some time now*.

Accordingly, we embark upon supplying an intellectual history to humanity's responsivity to existential risks. The aim is to reveal how contemporary X-risk research emerges from the broader sweep of human history. Our contention is that providing this edifying backdrop helps legitimise the furtherance of present initiatives.

This takes us to the Enlightenment. This period saw the consolidation of the various scientific vocabularies requisite for the first explicit prognoses on existential catastrophe. Yet the discovery of X-risk was a question of 'Enlightening', construed as humanity's global undertaking of self-responsibility, in an altogether more fundamental way. For, ultimately, it was only through realizing that we may *never reason again* that we became increasingly motivated to *reason ever better*, and, thus, were first summoned to the modernity-defining projects of long-term foresight, mitigation, and strategizing.

'The stakes are extremely large' (Baum et al., 2019, 28)'Here, then, is a very rational end of the world!' (Anonymous, 1816, 211)

1. Introduction

'Unless the human species lasts literally forever, it will some time cease to exist', writes Nick Bostrom. The sub-field of astrophysics titled physical eschatology establishes that, in cosmological timeframes, 'the long-term future of humanity is easy to describe: extinction' (Bostrom, 2009, 194). Intelligence, of whatever kind, will one day 'cease to exist' (Kraus & Starkman, 2000). Closer at hand, however, such certainty degrades into an ocean of near-term hazard. According to Torres (2017, 21), a 'growing swarm of risks' faces humanity as a planetary collective: ranging from artificial superintelligence to detonating supernovae, from weaponized pandemics to gamma-ray bursts. And yet, despite prevailing uncertainty, probabilistic reasonings from anthropic arguments—such as the Carter-Leslie Doomsday Argument (Leslie, 2002) or versions of SETI's Fermi Paradox (Ćirković, 2018)—insist that, whatever we currently judge our odds, they are likely significantly lower.

This expanding suite of human extinction threats or existential risks (hereafter X-risks) has, in other words, lately become the object of an emerging field of rigorous, quantitative, and scientifically serious inquiry. Baptized 'X-risk Studies' (Torres, 2017), the institutionalization of such research within bodies such as Oxford's Future of Humanity Institute should be understood as

* Corresponding author at: Oxford, UK.
E-mail address: thomas.d.moynihan@gmail.com.

https://doi.org/10.1016/j.futures.2019.102495
Received 29 April 2019; Received in revised form 3 December 2019; Accepted 14 December 2019
Available online 19 December 2019
0016-3287/

symptomatic of what Gramelsberger (2011, 20) has recently diagnosed as a widescale shift to the 'future perfect mode' in contemporary science. Demanding of us what Miller (2018) has called 'futures literacy', predictions are no longer used merely in hypothesis validation but are also perils—of increasing severity and scope—that must be prevented.

Across the board, we are ever more concerned with such risks. The growing edge of our present moment is progressively entangled with progressively distant futures. Appropriately, it was the German historian Koselleck (2004, 3) who specified modernization itself a heightening in the 'demands made of the future'. We now realize, however, that this is undeniably also an increase in the demands the future makes upon us. (This can be seen most immediately in anthropogenic climate change, which extends our horizon of moral culpability to indefinitely many future generations; elsewhere, one can also see this in what Bostrom calls the potential 'opportunity cost' of delaying space colonization efforts (Bostrom, 2003, 2013).) And yet, this tendency—of our increasing concern with our long-term fate as a species—itself has a history. Put differently, *we have been being swept up by the future for quite some time now*. This paper will recount this history, retelling the story of how we first came to care about the precarity of the human project.[1]

All the way back in 1836, the Italian pessimist Giacomo Leopardi proclaimed that, if *Homo sapiens* were to be extinguished, "the earth [wouldn't] feel that there is anything missing" (1982, 95). Three decades earlier, Marquis de Sade characteristically decreed that "nothing would be more desirable than the total extinction of humankind" (1968, 373). (He was duly arrested by Napoleon for such comments.) Earlier still, the influential French naturalist Comte de Buffon envisioned, in 1756, another lifeform inheriting our position of apex cogitator if the "human species were annihilated" (1797–1807, 27).

As ideas go, human extinction is a comparatively new one. It first emerged, during the period called the Enlightenment, spanned by the lives of the thinkers mentioned above. Yet, despite building preoccupation with the topic in recent years, it remains an idea without a history. Human extinction nonetheless has a history, and it is an important one, because it teaches us important lessons about what it means to be human in the first place. We mean this in the sense of what precisely is demanded of us by just such a status. For, to be a rational actor is to be a responsible actor, which involves becoming ever more responsive to the risks one faces. Therefore, retracing the story of how humanity first came to care about the risks it faces shows that such concern has always been of a piece with our rational undertaking of responsibility as a species.

Indeed, just as soon as the first predictions of existential catastrophe emerge during the Enlightenment, so too do the first projections upon plausible mitigative and strategic responses. These range from Lord Byron's 1824 picturing of a future humanity averting incoming comets by means of ballistic defence systems (Medwin, 1824, 226-8), to Benoît de Maillet's anticipation, as early as the 1720s, of the irrigation and rearranging of entire continents in attempts to offset the desiccating heat of an expanding sun in the deep future (de Maillet, 1750, 214-5). Elsewhere, the very first novel to tell the story of the 'Last Man' (written shortly before the author killed himself in 1805) envisioned a future wherein civilization has set titanic machines to work, levelling mountains and shifting seas, in order to extract diminishing nutrition from a collapsing biosphere in the far future (Grainville, 2002, 46-7).

Such grand and early visions of our macrostrategic resilience in the face of impending extinction alert us to the fact that the drama of how we first came to care about our precarity as a species is also the story of how humanity started to become responsible for itself. (An unfinished task, no doubt, given continued injustices and recklessness, but one that isn't illegitimate because of this.) For one is only responsible for oneself to the extent one understands the perils one faces and is driven to do something about them. This is suiting since the historical period in question, the Enlightenment of the eighteenth-century, explicitly defined itself along the lines of the human collective's undertaking of self-responsibility (Kant, 2013).

Why, then, be interested in such a history? Our contention is that recollecting the historical drama of how we first *came to care* for our extinction helps further establish why we must *continue to care* (and care now, as never before, inasmuch as the current century is, by many accounts, the riskiest yet (Rees, 2009)). It affirms that global risk management is the right thing to do moving forward. This is because such historical recollection shows today's efforts towards X-risk mitigation and macrostrategy as being the continuations and offshoots of a centuries-long project and calling of human self-improvement. Initiatives such as Oxford's FHI or Cambridge's CSER are here shown to emerge from a much wider and more expansive historical movement. In other words, retracing this history shows that such concerns have always been, and so remain, answers to the summons of Enlightenment. Casting contemporary efforts as emerging from out of this broad and progressive historical upswell edifies present-day research initiatives and additionally, we contend, helps justify them in the eyes of the present whilst also petitioning for their future advancement. This is because providing such a historical backdrop provides a novel argument for why present concern with global risk is ethically obligatory in that it reveals that such concern is presupposed by the very nature of rational agency itself: for insofar as we humans are capable of acting in accordance with the better reason, and are compelled to pursue the project of unceasing self-improvement that acknowledgment of this demands of us, we *cannot but* become concerned with existential risks and their mitigation. And, as we shall see, retracing the history of how we first came to think about X-risks confirms this to be the case. That is, it establishes that we first became able to even so much as think about human extinction through our ability to act and think rationally. And, by consequence, recollecting this history establishes that such concern unavoidably remains an essential and indispensable part of our vocation as a species of rational beings. (In this, we rely on the Enlightenment notion of a 'global human vocation' (Fichte, 1987), which consisted in the politically

[1] It must be acknowledged from the outset that the present paper focuses primarily on European and Enlightenment thinkers. However, the full story is properly global and intercultural. Take, for example, the fact that the mathematics essential to the later development of a probabilistic science of risk (notably the 'zero' required for place value notation and arithmetic) saw its first beginnings in medieval India; or, elsewhere, the fact that much of the requisite thinking on possibility and modal logic emerges from Arabic philosophers during the Islamic Golden Age. The later European developments covered in this article would not have been possible without these global crosspollinations.

momentous recognition that we are creatures that are—at least in part—accountable for what we think, what we do, and what it is that we are.) In other words, the story of how humanity started thinking about global risks reveals that caring about such threat has always been, and so remains, at the very core of what it even is *to be human* in the first place: the historical drama affirms that, inasmuch as we set ourselves the task of being rational, and inasmuch as we make ourselves worthy of such a title, we could not but have become concerned with such threats. It was always a core part of human enlightening and so it remains. This is why showing that concern for extinction emerged from such an edifying backdrop reaffirms that continuing concern remains similarly justified. In other words, it helps further establish that concern for our extinction has always been, and so remains, the *only rational action* moving forwards. (In this, we are greatly inspired by what the philosopher Robert Brandom calls 'recollective rationality' (Brandom, 2013, 2019).)

Ultimately, this is why we contend that supplying a historically reflective dimension to X-risk studies (and future studies more generally) may help further secure a notion of our 'global human vocation' that is fit-for-purpose, not only for our species to survive, but also to thrive. Such a history, retracing how our ability to even think about extinction emerges from our ability to follow 'what is right', helps reaffirm and re-establish that present concern for our extinction is *the right thing to do*. In line with this, one could call what we are here pursuing 'overt history as covert future-proofing'.

The article is structured along the following lines. First, Section 2 will explore how the idea of human extinction is conceptually, and genealogically, distinct from that of religious apocalypse; Section 3 will then explore some of the reasons why human extinction was not so much as even thinkable prior to a certain point in time. Thereafter, the article will explore three fields of empirical science—each of which first consolidated during the Enlightenment—which were requisite and necessary for oncoming understanding of existential risk. The first of these was the earth sciences (Section 4), the eighteenth-century appearance of which gave appreciation of natural history's vicissitudes; the second was demography and population science (Section 5), whose arrival inculcated first awareness of ourselves as a biophysical species; and the third was probabilism and actuary (Section 6), the emergence of which revealed our placement within a cosmic field of jeopardy and risk. To close the article, we explore how, though these empirical discoveries were essential, descriptions of observable fact were not alone sufficient to truly grasp human extinction. This is because, in order to truly appreciate the *moral stakes* involved in an extinction event (and thus become in the first place motivated to investigate this prospect as a grave tragedy), in-step reflections upon the placement of human value within a desacralized cosmos were also requisite. Namely, we needed to realise that our moral principles would not be part of the natural world independently of our ongoing upholding of them. Appreciating this could not come from describing brute facts alone, because it arrived from the discovery that our prescriptive norms are not derived from factual descriptions. In short, we had to first separate 'value' from 'fact' before we could become gripped by the potential fact of the end of all value. As such, alongside the three fields of empirical science explored, key philosophical and ethical breakthroughs were also necessary: to conclude, we explore these developments—which reached their culmination in the revolutionary work of Immanuel Kant—and connect them to our present situation.

2. Why extinction is not apocalypse

First, however, we must answer an obvious question. Where was the idea of human extinction prior to the Enlightenment? There has, of course, always been a tradition of apocalyptic prophecy and religious armageddon. Haven't people forever been worrying about such portents?

Extinction, bluntly, has nothing to do with apocalypse. It is different in kind as a concept. For a start, religious prophecies concerning apocalypse are designed to reveal the ultimate meaning and morality of things. (It's in the name: 'apocalypse' means 'revelation'.) Extinction, by direct contrast, reveals precisely nothing, and this is because it instead prognoses the end of morality and meaning itself. (If there are no humans, there is nothing humanly meaningful left.)

This, moreover, is precisely why extinction matters and is so ethically interesting. Judgement day allows us to feel comfortable knowing that, in the end, the universe is ultimately in tune with what we call justice (even if the sentencing is divine and inscrutable). In this way, nothing was ever truly at stake. Extinction, on the other hand, alerts us to the fact that all that we hold dear has forever been in jeopardy. Or, everything is at stake.

In short, apocalypse is premised upon a reification of our values and their identification with the independent universe; whereas extinction, by direct contrast, rests upon an appreciation that the universe is utterly unresponsive to our moral wishes and sense of justice.

Put differently, where apocalypse secures a sense of an ending, extinction prognoses the ending of sense. One is conciliatory, the other is inconsolable. As such, they are distinct and even incompatible conceptions. And as distinct in kind, they are also distinct in origin.

Yet cultural historians of the eighteenth-century have consistently missed the emergence of the idea of human extinction during the period. This is precisely because they haven't been clear on precisely this conceptual distinction between apocalypse and extinction. Clarke (1979, 43) is characteristic when he subsumes the first depictions of human extinction under a 'continuing mythology of doom'. They are, he claims a mere renewal of the 'immemorial fears' of 'archaic cosmogonies', extending from 'Ragnarok to Götterdämmerung', and are thus 'quarried' from the 'deeper levels of the psyche'. Though less overtly perennialist in outlook, many other commentators frame the matter as 'mythopoeic' (Stafford, 1994) or gloss it as 'secularized eschatology' (Wagar, 1982, 13).[2]

[2] Another popular line of argument is that the first imaginative works depicting human extinction were simply the products of the biographical traumas of the authors: that writers were merely abstracting their personal tragedies onto the human whole. Eva Horn, for example, has named this

The reason why this 'secular apocalypse' approach to the first speculations on human extinction largely fails as a historical explanation is precisely because of its ahistorical reliance on perennial and platonic archetypes (i.e. transhistorical mythologies 'quarried' from the 'deeper levels of the psyche'). In other words, it cannot answer *why* the idea emerged at a specific time—rather than any other—because it refuses itself the ability to specify what makes the idea distinctive and conceptually unique in the first place (by stressing its continuity with prior notions). For what makes human extinction special as an idea is that it marks a historical culmination of our awareness that human values are not rooted in independent nature nor would they persist without us being around to uphold them (whereas judgement day rests in the opposite presumption: that the universe has an inherent and indwelling sense of narrative justice). Apocalypse and extinction are, in this sense, false friends (in the sense borrowed from linguistics) in that, though their they may look similar at the level of aesthetic content (in their pyrotechny and tumultuousness, for example), they are premised, at a deeper conceptual level, upon incompatible notions of reality.

Again: where apocalypse ensures that the end of time has narrative meaning, extinction allows that all narrative and meaning may end within time. As such, having appreciated what makes the concept unique and distinctive (by suitably distinguishing it from the elder tradition of apocalypse), we can properly turn to the question of why predictions of extinction emerged at a specific point in time.

In other words, the notion of human extinction, requiring acknowledgement of our position as a biological species within a desacralized universe, was not so much as even thinkable prior to a certain point in history. Why was this?

3. Why human extinction was not a concern prior to the enlightenment

The answer to the riddle lies in a venerable philosophical assumption often dubbed the Principle of Plenitude. Persistent across Western intellectual history from Ancient Greece onward (Lovejoy, 1936), the Principle states that 'all legitimate possibilities are realised'. This entails that there can be no entirely unjustifiable absences in existence: no things that *could* have been, but simply just never are, without *any* further justification. This is why the prospect of an absence as seemingly unjustifiable as an extinction (inasmuch as it represents an unaccountable gap in nature's space of realizations) was unacceptable for most of Western history. As has been multiply documented, such an axiom remained a persistent theme of Western thinking from Aristotle down to Leibniz (Knuuttila, 1981).

Put simply, adherence to Plenitude long precluded comprehension of extinctions (whether human or non-human) because it forbids appreciation of the *irreversibility* requisite for such an event to be considered at all meaningful. Plenitude entails that, should any species be lost, the possibility of its returning will, eventually and inevitably, be fulfilled.

This prevented appreciation of species extinction from Antiquity to the Enlightenment (aside from one or two rare exceptions (Parejko, 2003)). Representative of this, Lucretius, in the first-century BC, confidently espoused that "nothing in creation is the only one" Lucretius(2007, 68). There can thus be no final instance of any natural kind: nothing can truly die out. Centuries later, in 1686, identical convictions led Bernard de ,Fontenelle (1687150-1), to declare that, even if our sun dies, "[n]o species [can] totally perish" because, across the cosmic infinities, all terrestrial species will inexorably return to repopulate some "New World".

Accordingly, one could prognose as many planet-shattering events as one liked, but, within the cosmos's "immense ocean of matter", nothing is ever truly lost and, as Denis Diderot (1966, 174) once put it, the "totality remains" pristine (and by "totality" Diderot meant the space of nature's realization of all possible species). Indeed, once asked whether humanity would one day go extinct at a dinner party, Diderot answered "yes", but immediately qualified this by saying that evolution would be re-run and "at the end of several hundreds of millions of years of I-don't-know-whats the biped animal who carries the name man" would inevitably return (Kors, 1976, 99).

In 1750, the British astronomer Thomas Wright, who elsewhere suggested that "the end of ye earth might [be] certainly predicted" by computing cometary paths (Wright, 1968, 32,), removed any moral stakes from such a forecast by assuaging that the "Catastrophy of a World, such as ours [and] even the total Dissolution of a System of Worlds" is, in light of the cosmos's eternal cycles of renovation, as insignificant as births and deaths upon our own planet (Wright, 1750, 76). Certainly, the overarching conviction that nature is as full of species as is possible led to a long-held assumption that every single planet is populated with beings just like us. Indeed, as Galileo had proclaimed in the 1630s, an uninhabited and unpopulated world is *naturally impossible* on account of it being *morally unjustifiable* (Galilei1661, 68). This, therefore, was why the "Dissolution of a System of Worlds" mattered little.

In line with this, a line from the poet Alexander Pope was often quoted and celebrated during the eighteenth-century. Talking of the immensity of space, Pope had, in 1734, equated the destruction of worlds to the trivialities of effervescing fizz. "And now a bubble burst, and now a world", he wrote (2016, 12). Only if you are confident that nothing can "totally perish" could such a statement be read—as indeed it was—as a sign of creation's jubilant magnanimity rather than its flippant malignance.

Confidence in the 'fullness' of nature's realization of each of its possible species even led Thomas Jefferson, at the close of the eighteenth-century, to argue that specimens such as the newly unearthed "mammoths or megalonyxes" must represent species still thriving and populous throughout the unexplored regions of the Americas (Jefferson, 1799). This, of course, was in the face of swiftly mounting evidence to the contrary. A few decades later, and after scientists could no longer deny that previous species had indeed disappeared, the same set of ideas nonetheless remained irresistible to Charles Lyell (the geologist who is often credited with

(*footnote continued*)
the 'poetics of extrapolation' (Horn, 2014, 72). However, legitimate though this argument may be, it is lacking as a *historical* explanation inasmuch as it cannot account for why predictions of human extinction did not occur to ill-fated individuals prior to the late eighteenth-century.

founding the field). In a book that would later become one of Darwin's favourites, Lyell declared that, even though they were indeed currently absent, prehistoric animals like iguanodons, ichthyosauri, and pterodactyls would, in some distant future, *return and reclaim the earth*. The disappearance of entire genera, Lyell proclaimed (2009, 164-5), designates a temporary "interval of quiescence" within nature. He spoke to colleagues of dinosaurs inevitably reclaiming the British countryside in the deep future (Rudwick, 1975, 558). Indeed, it was largely due to these kind of background beliefs that people did not even notice the Mauritian dodo's demise until the 1830s, even though it disappeared sometime around the 1690s (Turvey & Cheke, 2008). Theoretical assumption of Plenitude made extinction unavailable at the level of empirical observation.

Quite simply, long-standing belief in Plenitude blocked understanding of the *stakes* involved in extinction such that it was unworthy of observation (let alone forecast or strategizing). Whether human or animal, dying out was, after all, but a mere "interval of quiescence". For, if the guarantee of eventual return applies to dinosaurs, it applies also to humanity. Certainly, one contemporaneous commentator, responding to Lyell's ideas on the eternal return of species, jubilantly reached just such a conclusion: "Here, then, is no extinction for us!" (Nares, 1834, 240).

A related issue further obstructed our thinking on the matter. This was the much more general conviction that the cosmos itself is inherently imbued with value and justice. Again, this assumption dates back to the roots of Western philosophy. (Indeed, it was itself the motivation behind long-standing belief in Plenitude.) At the risk of over-generalization, pre-Enlightenment thinking is typified by a tendency to mingle human values and natural facts. Leibniz, as ever, offers the fullest elaboration of this long-standing position. He was unequivocal: "there is nothing fallow, nothing sterile, nothing dead in the universe" (Leibniz, 1991, 26). This was because anything truly "sterile" or "dead" would go against his conviction that it is the nature of nature to be morally just. We see this most clearly expressed in his famous notion that we live in the "best of all possible worlds", but similar tendencies are long ago exampled in notions like Plato's doctrine of the Idea, which taught that reality is intelligible precisely because it is essentially intellectual in its structure.

And, simply, where existence is presumed inherently rational, reason cannot itself cease to exist, such that the terminus of *Homo sapiens* can have no true sting nor stakes. This is why the Ancients and Medievals did not think about human extinction. The idea could not gain its full meaning as the irreversible end of moral activity and sapient values precisely because the cosmos itself was presumed to be imbued with such qualities. Again, existential catastrophe was not so much as even thinkable because the ethical stakes that pick it out as a concept were unavailable. In a universe inherently suffused with moral value, such a prospect is trivialized to the point of being unthinkable. In this pre-modern framework, the very *motivating occasion* behind the modern project of pre-empting and predicting the long-term future was absent. It was so axiologically trivial as to be unworthy of objective investigation or attention.

Correlatively, when Aeschylus (1947, 20-1) or Hesiod (2006, 103) long-ago wrote of Zeus's plan to "destroy this race [of] human beings", and when Plato (2004, 19) rehearsed this in his *Protagoras*, we observe that such episodes cannot denote an existential catastrophe in the sense now pursued by future studies. This is because such mythic episodes cannot comprise an end of sapient values. Even if humans are smitten by Zeus, human-like intelligence indefinitely lives on in the creator. Nearly 1500 years later, Aquinas again encapsulated such an outlook when he proclaimed that a world without any intellects within it is an "impossible supposition" (1952, 11). Put alternately, there is, for Aquinas, no possible world within which there are no minds.

Indeed, the notion that the intellectual structures that we use to categorize reality are not permanent features of independent existence itself only really begins to become available at the end of the medieval period. When, in 1623, Galilei (1957, 274) wrote that "if the living creature were removed, all these qualities would be wiped away and annihilated" he was expressing a comparatively novel notion.

And so, by 1770, Baron d'Holbach noted that long-running mingling of natural facts and mental values had historically prevented appreciation of the precarity of the human species. The idea that "*whatever is, is right*", he wrote, has long blocked the idea that "the human species is a production peculiar to our sphere" and, thus, is liable to "disappear". Assumption that the universe is full of value, he added, leads to the "conjecture that the other planets […] are inhabited by beings resembling ourselves" (Holbach, 1795-6, 146-7). In line with this, the prospect of our extinction could only become meaningful—and something that could begin to motivate anticipative efforts of prediction and pre-emption—when we realised that the universe was not itself brim-full with what we find valuable. Again, 'value' and 'fact' had to be disentangled before we came to be gripped by the prospective fact of the end of all values. Realizing this is what first made us worried for our future.

This realization, as we investigate below, took place during the eighteenth-century Enlightenment. It should be no surprise that this was the period when a notion of the future as radically open and uncertain first emerged (Koselleck, 2004). Indeed, the discovery of the future and the discovery of our extinction are not unconnected historical events: we first came to care for the long-term future precisely when we realised our place within it remains precarious. (Indeed, as Rescher (1998, 26) writes, only for 'those who take historical change seriously does the question of predicting the future become a matter of real concern'.) In other words, it was in acknowledging that what we think and do matters—existentially so—that we were first summoned to the project of protecting everything we find valuable and, thus, were motivated to predict our long-term future upon this planet.

In short, we had to first realize that intelligence is *astronomically precarious* before we could appreciate that it is *astronomically precious* and, thus, worthy of predictive and mitigative protection. Most immediately, first awareness of such precarity emerged, across the eighteenth-century, due to the appearance of three new-found areas of descriptive science. These were: geology, demography, and probabilism. To the first of these we now turn.

4. Enlightenment science #1: geoscience

London's *New Monthly Magazine*, in October 1816, published an article frenetically listing various extinction scenarios. The "aqueous fluid of our globe" is running out; the "thirty or forty thousandth deluge" is due; the "moon is to fall upon us". Elsewise, there is always the possibility of "general conflagration", or, "refrigeration of our globe". Notwithstanding the specific kill mechanism, we can be sure that in around "[f]ifteen hundred years" our "globe will not be habitable" and the "world will be at an end for us at least". And should non-human life survive all this, the Earth will—"billions of centuries" away—"tumble upon" the sun. This notwithstanding, all cosmic bodies are eventually destined to become "*caput mortuum*" (i.e. degenerated matter) in one "huge mass of dross". "Here, then, is a very rational end of the world!", the article closes. Though glibly concluding that "the amateur may take their choice" of their preferred extinction scenario, this article clearly displays what was then a new-found awareness of the instability of the earth system and of our precarious place within it.

This new appreciation of the vicissitudes of our planetary environ came from the emergence of the earth sciences during the eighteenth-century. Geology, that is, saw its first beginnings during the closing decades of the 1600s when two polymaths caused a stir by presenting some revolutionary theories to London's Royal Society. It was Robert Hooke and Edmond Halley who here proposed the first recognizably geohistorical conjectures. All this means is that they, for the first time, injected naturally caused vicissitude into their theories of how the earth works (previously, God was often presumed the cause of all major change or upheaval). Hooke proposed a terrestrial history rocked by gigantic earthquakes, wherein the planet's ever-changing surface led to "divers *Species* of Creatures" becoming "quite lost" (1996, 435). *This was the first ever unequivocal endorsement of the idea of species extinctions*. (Fossils have been documented since the Ancient Greeks (Mayor, 2000): however, due to background belief in the Principle of Plenitude no Ancient thinkers ever fully proposed the disappearance of species as truly irreversible.) Not long after Hooke's conjectures, Halley ventured the first image of something similar to what we now call a mass extinction event. He speculated, that is, that the "causal shock of a comet" could cause such a global cataclysm that "all things [living] should hereby be destroyed" (1723, 122).

The idea of prehistoric non-human extinctions thereafter became a topic of discussion amongst scientists. By the 1740s, Maupertuis (who elsewhere provided one of the first definitions of futurological "*prevision*" (Jouvenal, 1967, 12-3)) ventured that "the species we see today are but a small part of those [originally] produced" (Maupertuis, 1742, 24-6). During the 1760s, William Hunter used incipient comparative anatomical methods to argue that the recently uncovered Mastodon evidenced an organism no longer extant (Rolfe, 1985), and he later did the same with the Great Irish Elk. By this time, Voltaire (1765, 86-7) and his peers publicly endorsed prehistoric extinctions. During the 1770s, the German naturalist Petrus Camper turned his attention to the Woolly Rhino, thereafter drafting a paper on extinct quadrupeds (Meijer, 1999, 63-6). Around this time, eye-witness reports of the demise of contemporary species began returning to Europe from its colonies: ranging from Newfoundland's great auk (Roberts, 2007, 42) to the sea-cow of the Aleutian Islands (Sauer, 1802, 181). In 1793, in an influential paper, Carl Friedrich Kielmeyer (1993, 42–44) dwelt upon nature's profligate destruction of entire genera; and, by 1794, the first binomial classification had been granted to an extinct species of prehistoric cave bear (Rosendahl, Kempe, & Doppes, 2005).

Nonetheless, despite beginning to admit such mutability into their theories of nature, the prospect of human extinction remained largely absent from such writings. This, again, was due to background philosophical assumptions about the inherently moral design of the cosmos. Halley exemplified this: he justified his theory of global extinction events by saying that any future collapse is *morally justified* because it would rejuvenate the world for the benefit of some new returning civilization (he wrote about the destroying of our "whole Race for the Benefit of those that are to succeed" (Halley, 1723, 124)). In other words, Halley saw extinction as a morally justified truncation within a guaranteed cycle of returning human civilizations. (A 'punctuated eternalism', if you will.) Hooke held precisely the same position. Both scientists simply assumed presumed that an "annihilated" humanity would persistently reappear and repopulate the desolate earth after each global desecration.

By 1750, the French polymath surgeon Claude-Nicholas Le Cat repackaged this now-familiar cycle of "ruin and renovation" in his own geological theories. However, Le Cat was conspicuously unclear as to whether humans would, indeed, inevitably return after the next world-collapse. A shocked reviewer of Le Cat's book on the topic picked up on this equivocation, demanding to know whether "earth shall be re-peopled with new inhabitants" after any future desecration. In reply, Le Cat dodged the reviewer's accusation of departing from orthodoxy but not without wryly musing—with graveside smile—that there "are already a sufficient number of animals and men buried in the earth to gratify the curiosity of the new inhabitants of the new world, if there be any" (Anonymous, 1750, 384). The prevarication in the final clause surely did not appease Le Cat's perturbed reviewer. The cycle of the eternal return of species was slowly unwinding.

Indeed, Georges Buffon would soon insert undeniable temporal asymmetry and irreversibility into his theory of earth history. The influential naturalist theorized a future point of thermic death for our biosphere, triggered by the dissipation of all of the planet's internal heat, and he prognosed that this would take place 168,123 years hence (Haber, 1959, 118). When later broaching the topic of human extinction, Sade cited Buffon as an inspiration (de Sade, 1965, 332). Moreover, Buffon's peer, Jean-Sylvain Bailly, soon made the natural step of extrapolating this dissipative tendency toward "equilibrium" to *all cosmic bodies* (Brush, 1996, 77). Almost a century before Clausius named entropy in 1865, this was the first picturing of cosmic sterilization and extinction.

After Buffon's doomy forecast, it was another Frenchman, Georges Cuvier, who, in 1796, presented a paper providing 'irrefutable' empirical evidence 'for the reality of extinction' (Rudwick, 2008, 101). He did so via an application of comparative anatomy to Mammoth molars. Cuvier went on to formulate a theory of earth history riven with gigantic and planet-shuddering extinction events. No longer the stable cycle of return, this produced a new sense of the contingency of natural history and, thus, of our mutable place within it.

Unsurprisingly, it was not long after Cuvier was translated into English—in the early 1810s—that the first literary fictions

engaging the topic of human extinction appear. Appropriately, the years in which Cuvier was translated also saw one of the most catastrophic and colossal volcanic eruptions in human history: the detonation of Mount Tambora in Indonesia. Having injected megatons of sulphur into the stratosphere in 1815, Tambora's fallout caused a cascade of cataclysmic weather, harvest collapse, economic recession, cholera outbreak, subsistence crisis and geopolitical instability (D'Arcy, 2014). Almost blotting out the sun, 1816 became known as 'The Year Without a Summer'. This far-reaching geophysical catastrophe directly inspired a group of young authors to write the first proper literary engagements with human extinction in the English language. We refer to the literary circle of Lord Byron, Mary Shelley, and Percy Shelley. For, having been trapped indoors whilst holidaying in Switzerland due to titanic thunderstorms caused by Tambora's perturbations, the troupe began to discuss the topic of the longest-term prospects of humanity upon this unstable earth. Cuvier's theories, unsurprisingly, were in the background of these discussions (Brewer, 1994, 27–36). Appropriately, each of their next major works—following the *annus horribilis* of 1816—conspicuously engages the idea of human extinction.

Byron's 1816 poem 'Darkness' imagines what would happen to our planet if the sun were to suddenly go out. With chilling detail, it depicts the ensuing sterilization of our biosphere. By 1820, Percy Shelley's *Prometheus Unbound* contains a vision of fossilized prehistoric beasts, jumbled throughout the planet's crust, and dramatizes the threat of humanity joining this fossil pantheon. Most notably, Mary Shelley would go on, in 1826, to publish her *Last Man*. This was the first full-length novel depicting the truly global scope of an existential catastrophe, here at the hands of pandemic plague. Unsurprisingly, the influence of Cuvier's catastrophist geological theories can be seen in the background of each of these early experiments with the idea of human extinction (Bailes, 2015; O'Connor, 1991). Given this new appreciation of the vicissitudes of the earth, human extinction had become plausible: however, this plausibility was also resultant upon acknowledgement of our position as a biophysical species.

5. Enlightenment science #2: population science

Eight years before writing her *Last Man*, Mary Shelley had already alluded to human extinction in her early book on what is now called synthetic biology. We refer, of course, to *Frankenstein* of 1818. Therein, Shelley invoked the monster's potential, should Doctor Frankenstein make for it a female companion, to trigger humanity's extinction via outbreeding it as a competing germline (Shelley, 1999, 190–209). (The monster, indeed, is explicitly referred to as a "new species" (Shelley, 1999, 82).) This demographic threat of deadly population overshoot brings us to the second key science, also emerging during the Enlightenment, essential to the discovery of human extinction: the rise of so-called political arithmetic, or, demographic thinking.

Tellingly, one of the very first texts to engage in demographic ideas—written by the political thinker Baron de Montesquieu in 1721—was likewise one of the very first to mention the natural plausibility of human extinction. That is, in his *Lettres persanes*, Montesquieu declares that global population has diminished since Antiquity. "After doing calculations as exact as possible", he writes that he was ascertained that our "population continues to diminish daily, and if this trend persists within ten centuries the earth will be nothing but an uninhabited desert" (Montesquieu, 2008, 150). He was, of course, utterly wrong in this forecast. Nonetheless, as is betrayed by his use of words like "calculations" and "trend", Montesquieu is here employing the then new-found understanding that numbers can be applied to reality as 'time-stepping' procedures in order to *predict* its long-term future course (Gramelsberger, 2011). Demography, indeed, was one of the first sciences to put this idea into practice.

It is no accident, therefore, that multiple demographic treatises across the ensuing century contained similar musings on the future prospect of human cessation. In 1754, David Hume penned an essay on "populousness", responding directly to Montesquieu. Here, Hume (2007, 108) also pronounces that *Homo sapiens*, just like all other species, will eventually undergo extinction. Decades later, the political thinker Godwin (1798, 452-3), otherwise an ardent optimist in matters of political arithmetic, wrote that "[t]he globe we inhabit bears strong marks of convulsion", which natural scientists "agree to predict will one day destroy the inhabitants of the earth". In a later text, Godwin recorded the fact that the burgeoning field of population science had—across the previous century since Montesquieu—provided a rich arena for prognostications upon "the extinction of our species". Here, from 1820, Godwin looked back to Montesquieu's original forecast of 1721, and recalled the latter's grim prediction that "the human species is hastening fast to extinction" (Godwin, 1820, 100).

Demography initially arose from applications of nascent probability theory to mortality rates in the 1600s. This was pursued in order to attempt to compute annuity payments. It was as a by-product of such endeavours that inquirers first noticed that what we now call 'population' is an object in its own right: with its own tractable regularities, lawlike features, and dispositional properties. Only after these first steps in this new actuarial science (such as ,Graunt's 1662 discovery of statistical regularities in death rates) did the previously invisible item called 'population' first solidify as a target of objective investigation (insofar as its dynamics suddenly became capable of being numerically retrodicted and predicted). Thus, as the microscope was to the bacterium—and the telescope to the stars—so statistics was to our global human mass.

With this concrescence of population as a tractable scientific object, there, of course, came new avenues of power (Foucault, 2007, 75). But there also came a new unit of potential perishing. In other words, conceiving of humanity as a planetary collective was requisite for us to first become concerned with our precariousness at this scale. And so, ironically, computations of riskiness that had first made population visible to us simultaneously ensured that it was now subject to risk.

Indeed, aside from wielding the new-found understanding that numbers can be applied to reality in order to predict its future, the rise of demography was a crucial factor in the discovery of human extinction because it cemented humanity's awareness of itself *as a biological species*. For following John Ray's biological work in the 1680s, the idea of "species" had itself become scientifically defined as an organic form fixed, across time, by sexual propagation (Ray, 1686). Accordingly, through thereafter focusing attention upon humanity as itself a reproductive community, demography inculcated in us 'taxonomic self-awareness', or, appreciation that we are ourselves a "species". This was consecrated in Carl Linnaeus's inclusion, in 1758, of the genus "*Homo*" in his taxonomic system

(Linnaeus, 1758). Indeed, it was during this century that we first began referring to ourselves *as* "the human species" (Foucault, 2007, 75). And, simply, with taxonomic self-awareness comes consequent awareness of taxonomic precariousness. Or, thinking of ourselves as a species, we became able to think about our dying out as a species. (Moreover, arithmetical thinking on the matter highlighted the numerically granular nature of species decline: this caused first appreciation that we could die out in a stepwise and asynchronous manner, rather than simultaneously as apocalypse invariably imagines, this was key to ensuing narratives on the 'Last Man'.)

Certainly, all the way back in 1721, Montesquieu (2008, 151) had already catalogued a dizzying surfeit of disasters that have brought humanity "within a hair's breadth of extinction". He declared that any number of factors "may be at play" that can decimate our number at any moment. (Such statements were soon confirmed in the ruins of Lisbon, which was utterly decimated in 1755 by an unforgiving troika of earthquake, tsunami and fire.) This brings us to the third new field of science that was necessary for us to become gripped by our own extinction. Essential to the emergence of demography, it was the prior consolidation of a mathematized understanding of risk, probability, and uncertainty. Indeed, Montesquieu wrote these statements three years after the first probability textbook was published (de Moivre, 1718).

6. Enlightenment science #3: probabilism & actuary

Risk was conceptually formalized *post hoc*, undergoing an intensely belated mathematical birth in seventeenth-century Europe. And yet, it retroactively commands a vast 'prehistory'. By this we mean that anticipation of hazard is, of course, the universal backdrop of sentient existence. However, specifically *self-conscious* representation of oncoming dangers is possibly uniquely human. This is likely derived from language-use, and the fact that any language functionally presupposes the ability to talk about the permissible and impermissible and, thus, also the possible and impossible. This allows the ability to mentally simulate the future and self-consciously manipulate such simulations. Such an endowment has been called 'proscopic chronesthesia' (Tulving, 2002) or, more prosaically, 'mental time travel' (Suddendorf & Corballis, 2007). It explains why we appear uniquely able to anticipation non-present, unseen, and unexampled threats across arbitrarily large spatiotemporal distances. It also explains why we are able to manipulate and design our responses to oncoming dangers, relying on planning and strategizing rather than inherited or instinctual fear-response. Such a skillset explains why *only* humans developed agriculture and urbanized: advances representing our first widescale, albeit *ad hoc*, attempts at risk mitigation and risk distribution.

Aside from piecemeal advances throughout the ages in crop specialization and city defences (Merchant, 2016, 64-7), the frontier of human risk response remained constrained to mere emendations of such extemporaneous buffers all the way down to the Renaissance. Then, in the 1600s, something changed. This century saw the entrance of the word "risk" itself into the English lexicon, spreading first through the argot of maritime traders and their underwriters (Luhmann, 1996). During this period, with the contemporary explosion of insurance industries alongside inception of financial markets and speculation thereof, possible futures quickly became profitable and jeopardy a lucrative business. Indeed, financial derivatives became morally maligned in so-called Tulipomania (the first recorded speculative economic bubble) *just as* governments began funding themselves by selling annuities. And so, as the 'Art of Conjecture' dawned, risk became mathematically tractable and predicted futures increasingly grew to influence the present.

All this began in c.1552, when Gerolamo Cardano first formalized 'Games of Chance'. His breakthrough was in enumerating an abstract sample space for the die and interpreting each dice throw as an expression of this wider state space. He then proceeded to use numeral notations to track frequencies within this reference class (Cardano, 2015). With this first formalization of probability, Cardano combined our evolutionary endowment of anticipation with the rigorizations of mathematics, thereby fomenting modernity's love affair with riskiness and the sciences of decision. This, when compounded with the seventeenth-century's inception of calculus (and its ground-breaking ability to predictively model continuous natural processes), set the seeds for our current-day apparatus of planetary forecast and our ongoing tendency to become progressively swept up by the future.

Cardano's breakthroughs would have to wait until 1654, however, to be applied to predictive forecast proper. This was achieved in Pascal's celebrated solution to the so-called 'Problem of the Points' within his correspondence with Fermat. Here, in calculating the odds of who *would have* won an unfinished game, numbers were deployed, for the first time, to robustly weight future outcomes (Devlin, 2010).

From here, it was appropriately not long until probabilism was leveraged to compute the odds of what we now call global catastrophic risk. For, following Pascal, mathematicians like Jakob Bernoulli, in the early 1700s, extracted probability calculus from the gambling table and demonstrated its applicability to real-world affairs (Gorrochurn, 2016, xxi). It was, accordingly, Jérôme Lalande, who, in 1773, first applied probability to the question of existential threat. The French astronomer calculated the odds of Earth's deadly intersection with a comet as "*76 mille contre un*", or, 1/76,000 (Lalande, 1773, 30). This provoked panic on the streets of Paris due to sensationalized reporting in newspapers (Stewart, 1986). Nonetheless, we have here *the very first probabilistic forecast of an X-risk*.

Shortly after, du Séjour extended Lalande's work (du Séjour, 1775) and, not long after this, the ingenious Pierre-Simon Laplace became excited by such calculations (Hahn, 2005, 68). Laplace went on to compute his own odds of an impactor event. Despite vacillating on precise likelihoods, he famously maintained that "the small probability of this circumstance may, by accumulating during a long succession of ages, become very great". He imagined to himself what such an impact would look like: predicting drastic alteration of the earth's axis, shifting oceans, multiple species wiped out (Laplace, 1809, 63-4). Indeed, the article on a "very rational end of the world", cited earlier, refers directly to Laplace's and Lalande's computations on this matter: the "probability of such a disaster is daily increasing" the article facetiously notes (Anonymous, 1816).

And so, by 1810, the German astronomer Wilhelm Olbers converted Laplace's "long succession of ages" into a precise timeframe. He computed a stretch of 220 million years per collision (Olbers, 1810). (He was under-estimating, however: contemporary

calculations put serious collisions, capable of causing deadly impact winters, at once every 500,000 years (Napier, 2008, 226).) Just like Laplace's and Lalande's before him, Olbers's 'cosmic risk analysis' was numerously reported on and it reverberated throughout the popular press (Anonymous, 1819). With yet more scientists producing their own cometary forecasts and impactor probabilities thereafter (Arago, 1832, 48; Milne, 1828, 116), we see how it was that the unveiling of nature's aleatory dynamics helped facilitate a fundamental reconfiguration of our relationship to it. Our cosmic backdrop was no longer considered a cradle of infinite moral worth and eternally returning humanoids but was instead shown to be an enveloping field of roaming hostilities and risks. Indeed, where Galileo had, in the 1660s, proclaimed an unpopulated world naturally impossible on account of it being morally unaccountable, Olbers, in 1802, theorized that the Mars-Jupiter belt constitutes the desolate ruins of a shattered planet (Zach, 1802). Such a planet-sized gap in nature's space of realisations simply could not be morally justified or accounted for. The old cosmological insurance scheme was breaking down.

Alongside assessing the objective likelihood of planetary catastrophes another usage of probability had long been implied: as opposed to measuring the objective frequencies of events observed, this alternate conception centred upon the strategic position of the observer herself (Daston, 1995, 226-95). For, ever since 1662's *Port-Royal Logic* first applied probabilities to the mechanics of inference and Pascal's ensuing wager of 1670 conducted risk-benefit analyses concerning religious belief, there had been insinuation of this alternate guise of probability (Hacking, 1975). Later dubbed subjective probability (Carnap, 1950; Daston, 1994), it interpreted probability not as an objective frequency but as a degree of credence in a subjective belief. This is important because the traditional idea of objective probability necessarily bounds indecision and uncertainty to variability within an observed reference class; in this way it inherently restrains 'threat' to hazards previously experienced and witnessed. This is of key importance concerning thought upon human extinction because 'absolutely destructive events, which humanity has no chance of surviving […] completely annihilate our confidence in predicting from past occurrences' (Ćirković, 2008, 123). So it was that, in the 1760s, Reverend Bayes (1763, 392-3) wrote of our need for a rule of inference where we "absolutely know nothing antecedent to any trials". Bayes was, of course, responding to the academical problem of "inverse probability" and not thinking about existential peril. Yet, nonetheless, Bayes bequeathed to later ages a theorem that could be wielded to reason rigorously upon those risks that are known, in our own time, as 'unknown unknowns' or 'wild card risks'. Basically, as a gauge and measure of our ignorance, Bayesian subjective probability allows us to take into account the strategic position of observation itself and, thus, informatively reason upon the fact that our own ignorance concerning 'unknown unknowns' can itself be a measurable threat. In short, risk was no longer exclusively something observed, but something that subsumes observation itself.

7. Enlightenment philosophy: reason, risk & responsibility

And so, given new awareness of the vicissitudes of earth's history, of our precarious position within it as a biophysical species, and of our wider placement against a cosmic backdrop of jeopardy, we were finally in a position to become receptive to the prospect of our extinction. Yet, as explored in Section 3, none of this could truly *matter* until 'fact' was fully separated from 'value'. Otherwise, one could simply fall back on the age-old presumption that other planets are populated with humanoids like us or that beings like us will simply re-emerge (because it is the universe's nature to be as full of worth as is possible). Indeed, even Laplace would qualify his forecast of cometary catastrophe by assuring that civilization would simply later re-emerge. Everything, he assuaged, would "be done again" (Laplace, 1809, 63-4).

Accordingly, the final piece of the puzzle came, therefore, not from empirical science but from philosophy. It came from growing awareness of the distinction between prescription and description, or 'ought' and 'is'.

That is, a key driver behind the philosophy of the Enlightenment was growing realisation that moral values are a question of self-legislation. This master idea of the Enlightenment reached its culmination in the philosophy of Immanuel Kant. Indeed, Kant realised that values are maxims that we elect to bind ourselves by and are, accordingly, utterly dependent upon this ongoing election. Correlatively, such values should not at all be considered part of the furniture of the independent natural world apart from our active upholding and championship of them. This entails that our values are entirely *our own responsibility*, or, we are culpable for everything that we care about and cherish. In other words, given that the maxims that we champion would not be persistent features of the natural world outside of our ongoing stewardship of them, they thereby also demand our vigilant guardianship. We simply cannot rest assured believing in eternally returning humanoids or any other such naïve trivialization of the stakes involved in what we think and do.

Coming to realise this is what first summoned us to the modernity-defining projects of prediction, pre-emption, and strategizing. For, again, Kant explicitly defined 'Enlightenment' as humanity's progressive undertaking of responsibility for itself; and one is responsible for oneself to the exact extent that one is aware of the risks one faces and is moved to do something about them. This dynamic is what drags—and continues to drag—our theoretical and practical concerns further and further into futurity. Accordingly, our responsivity to the risks that face us as a species and at ever greater scales has always been, and so remains, part and parcel of the Kantian project of Enlightening. For in realising that 'minded agents' are entirely responsible for the entire fate of 'mind' we first answered the calling to the task of pre-emptively protecting and redoubting intelligence within an otherwise silent and uncaring universe.

In short, part of being a rational animal is therefore becoming concerned for the extinction of rationality. A rational being *cannot but* come to care for such prospects inasmuch as to be a rational species is to be a responsible species. It is merely a question of following the 'better reason' and that is all that it is to be rational. Accordingly, it is only suitable that, as he aged, Kant himself became progressively occupied with portents of existential threat.

In his immature work of the 1750s, Kant unreflectively endorsed Plenitude: going so far as to write that "we ought not lament the

perishing of a world as a real loss of Nature". Picturing what he called the "Phoenix of nature, which burns itself only [to] revive again" again and again throughout "infinity", he declared that "Man [is] himself not excepted from this law" of cyclical return. As was usual at the time, he then cited Pope's comparison of shattered worlds to mere bubbles burst (Kant, 1900, 150). Nonetheless, upon entering his mature phase in the 1780s, Kant started to become increasingly concerned with the prospect of human extinction. Starting from his third critique's reference to "nature's most ancient revolutions" in 1790 (Kant, 1987 305), the spectre of extinction emerges again and again. By the writing of his *Anthropologie*, Kant imagined primates becoming sapient and replacing humanity (2006, 232-3). In 1795's essay on *Perpetual Peace*, Kant gloomily proclaims that the titular "perpetual peace" may only be achieved in the "vast graveyard of the human race" (1991, 96). Thereafter, in 1798, in an essay on the methodology of predicting the long-term future, Kant taxonomized plausible long-term trajectories for humankind. (His taxonomy of collapse, perfectibility, and recurrent oscillation nicely resembles Bostrom's own taxonomy of the same (Bostrom, 2009).) Here, musing whether human history will be indefinite, Kant is arrested by the plausibility of some global disaster "that will push aside the human race to clear the stage for other creatures" (1979, 161). This threat would again appear in his final work: here he worries that we might be replaced by some descendent species that, being more intelligent than us, extinguishes us and replaces us as the planet's apex cogitator (1993, 66-7).

8. Conclusion

In 1781, Kant (2007, 648) related our practices in everyday reasoning to a game of "betting". People often "pronounce their views" with seeming certainty, he noticed, yet they actually value their "persuasion [at] one ducat, but not at ten". Certitude, in other words, always comes in degrees. And yet, this is because incertitude—or jeopardy—is the very avenue through which we reach better, more accurate, claims. This was, at the time of Kant's writing, a revolutionary and auspicious notion: reason is rational, not because it provides certainty, but because it is self-correcting; or, not because it offers inviolable foundations, but because it is relentlessly revisable. Indeed, as Bayesianism first dimly implied at around the same time (cf. Swanson, 2016), it is by submitting our beliefs to the continual risk of invalidation that we update invalid beliefs and thus arrive at ever better ones. Only through this unceasing jeopardization of our beliefs can we truly say that we are responsible for what we think and do (because one is only accountable for one's commitments to the extent that one is willing to correct them if they are proven wrong). Jeopardy is, in this way, the very medium for the making and staking of ever more objective and justified claims. *Risk, reason, and responsibility are all intimately entwined.*

This was a brand-new realisation, dawning during the Enlightenment. It is important for our purposes because it encapsulates a watershed moment within a larger shift taking place across human history: a shift from thinking of the human mind and its values as being cradled—with absolute security—within a perfectly hospitable universe, toward a later acknowledgment that our cosmic backdrop is not a gigantic life-support system but is an enveloping field of risks, challenges, and hazards. And yet, though this development might seem solemn or alienating, it is *only* by outgrowing all such illusions of secure foundations (and their related 'existential insurance schemes'—such as Plenitude) that we answer our calling as rational animals (as is obscurely implied in Kant's comments on the relation between reasoning and gambling). This truth is scalable from the level of an individual's inference up to that of the species as a whole: for it was only through articulating the risks involved in what we think and do, and doing so on increasingly encompassing scales, that we were progressively summoned to the task of undertaking responsibility for ourselves and our position upon this planet, so as to become ever more worthy, in our ongoing practice, of the title 'rational'.

In line with this, we note that the historical discovery of human extinction was itself a summons to continual self-betterment. For, though sombre, realising we could die out was a crowning achievement of Enlightenment: in that, ever since this awakening, we realise that we must act and reason ever better because, should we not, we may never reason ever again. This is why concern for human extinction cannot but be anything other than rational. The history shows that it is in fact central to our image of ourselves as creatures endowed with rationality. For, ultimately it is through reflection on existential risk that we realise that minded agents are responsible for the entire fate of mind—whether it will have been inglorious extinction or Kardashev scale flourishing (Kardashev, 1984)—because it is only through invoking such universalized culpability that we are in the first place motivated and compelled toward ever better prediction, pre-emption, and strategizing, so as to face, as a collective, the challenges of a progressively riskier world. (Where here, once again, acknowledgement of jeopardy is revealed to be the very avenue of our improvement.)

Accepting this, it becomes clear that, across history, our growing responsivity to hazards of increasing severity and scope has always been part and parcel of the undertaking we still call 'Enlightening'. Starting from Kant's otherwise unassuming comments on the relation between reasoning and risk-taking, this dynamic has snowballed, in our own time, into scientific research into macro-strategy, existential perils, and planetary-scale mitigation: daring pursuits which can accordingly be cast as the crest of a wave that began multiple centuries ago.

It is useful to see this frontier of modern inquiry with its full historical backdrop in tow. For, as mentioned, to give something a history of this sort is to legitimize it in the eyes of the present. More so, it is also to petition for its future advancement and furtherance. Retracing this history, in other words, positions today's emerging science of global risk (as championed by current initiatives from Oxford's FHI to Cambridge's CSER) as the fruition and apex of centuries of hard work and human progress. As mentioned, this is edifying. Yet it needn't be edification alone. For to unveil existential risk mitigation as a climax of the much wider and more expansive undertaking of human self-betterment (an undertaking that, by definition, remains unfinished and incomplete) is to foment the furtherance of these initiatives in the present. This is because the history shows that concern for existential risk was always an indispensable component of our coming to answer our calling as rational beings who are accountable for themselves and their values, and yet this calling is an undertaking that necessarily always remains incomplete and unfinished such that recollecting how and why we first came to care helps reaffirm why we must continue to care. This is because, simply, this recollection reveals

investigation into global risks to have always been an essential and indispensable component of what it even is *to be human* in the first place (inasmuch as the human has come to define itself as the rational, and thus responsible, being). Historically speaking, then, this establishes that concern for X-risk is at the very core of our species' vocation. And so too must it remain central to our project going forward. For, again, a rational agent—insofar as it is rational—*cannot but* come to spell out the existential stakes involved in what it thinks and does within the world. It is the very duty of rationality to do so. In other words, it is only in recognising the risks involved in our enterprise that we are initially summoned to such a daring vocation of self-assertion and self-betterment; and, in this sense, recollecting the history of how we first answered this calling, or set for ourselves such a task, is simultaneously the summons to persevere with it, come what may. For the ongoing project of human betterment is an undertaking that we inherit from our collective past (in that no one would be able to so much as acknowledge such a calling shorn of history's tutelage) but it is also an undertaking that, by this same token, remains incomplete and unfinished (because one isn't truly following 'the good' if one isn't also pursuing 'the better' so as to supersede the mishaps of the past). And this is why it is useful, in light of the existential threats that loom on our collective horizon, to once again champion the Enlightenment notion of humanity as a daring enterprise or vocation: in that the human project is something that, across the centuries, we have come to acknowledge can either *succeed* or *fail* and the more we accept responsibility for this truth the more we are compelled to do what is righteous, rational and responsible within the world. Looking at our current situation and its perils in light of this sweeping movement of history gives us warrant for hope for our future on this planet and, possibly, beyond.

Acknowledgment

I was funded by the United Kingdom's AHRC (Arts and Humanities Research Council) from 2014-2017.As such.This work was supported by the UK Arts and Humanities Research Council.

References

Aeschylus (1947). *Prometheus bound (R. Warner, trans.).* London: Bodley Head.
Anonymous (1750). An account of several systems, particularly that of the ingenious Mr. Le Cat. *The Monthly Review, 3*, 444–459.
Anonymous (1816). Of the end of the world. *The New Monthly and Universal Register, 6*(33), 209–211.
Anonymous (1819). Antiquarian and philosophical researches. *The Gentleman's Magazine, 89*, 541–542.
Aquinas, T. (1952). *The disputed questions on truth (Vol. 1) (R.W. Mulligan, trans.).* Chicago: Regnery.
Arago, F. (1832). *Tract on Comets: And particularly on the Comet that is to intersect the Earth's Path in October, 1832 (J. Farrar, trans.).* London.
Bailes, M. (2015). The psychologization of geological catastrophe in Mary shelley's The Last Man. *English Literary History, 82*(2), 671–699. https://doi.org/10.1353/elh.2015.0018\.
Baum, S. D., Armstrong, S., Ekenstedt, T., Häggström, O., Hanson, R., & Kuhlemann, K. (2019). Long-term trajectories of human civilization. *Foresight, 21*(1), 53–83. https://doi.org/10.1108/FS-04-2018-0037.
Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London, 52*, 370–418. https://doi.org/10.1098/rstl.1763.0053.
Bostrom, N. (2003). Astronomical waste: The opportunity cost of delayed technological development. *Utilitas, 15*(3), 303–314. https://doi.org/10.1017/S0953820800004076.
Bostrom, N. (2009). The future of humanity. In J. B. Olsen, E. Selinger, & S. Riis (Eds.). *New waves in philosophy of technology* (pp. 186–216). New York: Palgrave Macmillan.
Bostrom, N. (2013). Existential risk prevention as global priority. *Global Policy, 4*(1), 15–31. https://doi.org/10.1111/1758-5899.12002.
Brandom, R. (2013). *Reason in philosophy: Animating ideas.* Massachusetts: Harvard University Press.
Brandom, R. (2019). *A spirit of trust: A reading of Hegel's phenomenology.* Massachusetts: Harvard University Press.
Brewer, W. D. (1994). *The Shelley-Byron conversation.* Florida: University of Florida Press.
Brush, S. (1996). *A history of modern planetary physics: Nebulous earth, the origin of the solar system, and the core of the earth from Laplace to Jeffreys.* Cambridge: Cambridge University Press.
Buffon, G. L. L. (1797–1807). Natural History (Vol. 6). (J.S. Barr, trans.) London.
Cardano, G. (2015). *The book on games of chance: The 16th-Century treatise on probability (S.H. Gould, trans.).* New York: Dover Publications.
Carnap, R. (1950). *Logical foundations of probability.* Chicago: University of Chicago Press.
Ćirković, M. M. (2008). Observation selection effects and global catastrophic risks. In N. Bostrom, & M. M. Ćirković (Eds.). *Global catastrophic risks* (pp. 120–145). Oxford: Oxford University Press.
Ćirković, M. M. (2018). *The great silence: Science and philosophy of Fermi's paradox.* Oxford: Oxford University Press.
Clarke, I. F. (1979). *The pattern of expectation, 1644–2001.* New York: Basic Books.
D'Arcy, W. (2014). *Tambora: The eruption that changed the world.* Princeton: Princeton University Press.
Daston, L. (1994). How probabilities came to be objective and subjective. *Historia Mathematica, 21*(3), 330–344. https://doi.org/10.1006/hmat.1994.1028.
Daston, L. (1995). *Classical probability in the enlightenment.* Princeton: Princeton University Press.
Devlin, K. (2010). *The unfinished game: Pascal, fermat, and the seventeenth-century letter that made the world modern.* New York: Basic Books.
Diderot, D. (1966). *Ramaeu's Nephew/D'Alembert's Dream (L. Tannock, trans.).* London: Penguin.
Fichte, J. G. (1987). *The vocation of man (P. Preuss, trans.).* Indianapolis: Hackett.
Fontenelle, B. (1687). *A plurality of worlds (J. Glanvill trans.).* London.
Foucault, M. (2007). *Security, territory, population: Lectures at the college de France, 1977-8 (G. Burchell, trans.).* New York: Palgrave Macmillan.
Galilei, G. (1661). *Mathematical collections and translations (Vol. 1) (T. Salisbury trans.).* London.
Galilei, G. (1957). *Discoveries and opinions of Galileo (S. Drake trans.).* New York.
Godwin, W. (1798). *Enquiry concerning political justice, Vol. 1* London.
Godwin, W. (1820). *Of population: An enquiry concerning the power of increase in the numbers of mankind, being an answer to Mr. Malthus's essay on the subject.* London.
Gorrochurn, P. (2016). *Classic topics on the history of modern mathematical statistics: From laplace to more recent times.* New York: John Wiley & Sons.
Grainville, J.-B. F. X. C. (2002). *The last man: New English translation (B. Stapleford, trans.).* Connecticut: Wesleyan University Press.
From science to computational sciences: A science history and philosophy overview. In G. Gramelsberger (Ed.). *From science to computational sciences: Studies in the history of computing and its influence on today's sciences* (pp. 19–44). Zurich: Diaphanes.
Graunt, J. (1662). *Observations made upon the bills of mortality.* London.
Haber, F. C. (1959). *The age of the world: Moses to Darwin.* Baltimore: Johns Hopkins Press.
Hacking, I. (1975). *The emergence of probability: A philosophical study of early ideas about probability, induction, and statistical inference.* Cambridge: Cambridge University

Press.
Hahn, R. (2005). *Pierre Simon Laplace 1749–1827; A determined science.* Massachusetts: Harvard University Press.
Halley, E. (1723). Some considerations about the cause of the universal deluge. *Philosophical Transactions of the Royal Society of London, 31*, 118–126. https://doi.org/10.1098/rstl.1724.0023.
Hesiod (2006). *Theogony, works and days, testimonia (G.W. most, trans.).* Massachusetts: Harvard University Press.
Holbach, P.T. (1795-6). The System of Nature (Vol. 1) (W. Hodgson trans.). London.
Hooke, R. (1996). *Restless genius: Robert Hooke and his earthly thoughts.* Oxford: Oxford University Press.
Horn, E. (2014). The last man: The birth of modern apocalypse in Jean Paul, John Martin, and Lord Byron. In N. Lebovic, & A. Killen (Eds.). *Catastrophes: A History and theory of an operative concept*. Berlin: Gruyter.
Hume, D. (2007). On the populousness of ancient nations. In E. Rotwin (Ed.). *Writings on economics*. Brunswick: Transaction Publishers.
Jefferson, T. (1799). *A memoir on the Discovery of certain bones of a clawed kind in the Western part of Virginia. Transactions of the American philosophical society, Vol. 4*, 246–260. https://doi.org/10.2307/1005103.
Jouvenal, B. (1967). *The art of conjecture (N. Lary, trans.)*. New York: Basic Books.
Kant, I. (1900). *Kant's cosmogony: As in his essay on the retardation of the rotation of the Earth and his natural history and theory of the heavens (W. Hastie, trans.)*. Glasgow: James Maclehose & Sons.
Kant, I. (1979). *The conflict of the faculties (M.J. Gregor, trans.)*. New York: Abaris Books.
Kant, I. (1987). *The critique of judgement (W.S. Pluhar, trans.)*. Indianapolis: Hackett Publishing.
Kant, I. (1991). *Political writings (H.B. Nisbet, trans.)*. Cambridge: Cambridge University Press.
Kant, I. (2006). *Anthropology from a pragmatic point of view (R.B. Louden, trans)*. Cambridge: Cambridge University Press.
Kant, I. (2007). *The critique of pure reason (M. Müller trans.)*. London: Penguin.
Kant, I. (2013). *An answer to the question: What is enlightenment? (H.B. Nisbet, trans.)*. London: Penguin Books.
Kardashev, N. (1984). On the inevitability and the possible structures of supercivilizations. In M. D. Papagiannis (Ed.). *The search for extraterrestrial life: Recent developments* (pp. 497–504). Reidel: Dordrecht. https://doi.org/10.1007/978-94-009-5462-5_65.
Kielmeyer, C. F. (1993). *Über die Verhältniße der organischen Kräfte unter einander in der Reihe der verschiedenen Organisationen, die Gezetze und Folgen dieser Verhältniße.* Marburg: Basilisken Presse.
Knuuttila, S. (Ed.). (1981). *Reforging the great chain of being: Studies in the history of modal theories*. Reidel: Dordrecht.
Kors, A. C. (1976). *D'Holbach's coterie: An enlightenment in Paris.* Princeton: Princeton University Press.
Koselleck, R. (2004). *Futures past: On the semantics of historical time. (Keith Tribe, trans.)*. New York: Columbia University Press.
Kraus, L. M., & Starkman, G. D. (2000). Life, the universe, and nothing: Life and death in an ever-expanding universe. *Astrophysics Journal, 531*(1), 22–30. https://doi.org/10.1086/308434.
Lalande, J. (1773). *Réflexions sur les comètes qui peuvent approcher de la terre.* Paris.
Laplace, P.-S. (1809). *The System of the World (Vol. 2). (J. Pond, trans.)*. London.
Leibniz, G. W. (1991). *Monadology (N. Rescher, trans.)*. Pittsburgh: University of Pittsburgh Press.
Leopardi, G. (1982). *Operette Morali: Essays and dialogues (J. Galassi, trans.)*. Berkeley: University of California Press.
Leslie, J. (2002). *The end of the world: The science and ethics of human extinction.* London: Routledge.
Linnaeus, C. (1758). *Systema naturae.* Stockholm.
Lovejoy, A. (1936). *The great chain of being: A study in the history of an idea.* Massachusetts: Harvard University Press.
Lucretius (2007). *The nature of things (A.E. Stallings, trans.)*. London: Penguin.
Luhmann, N. (1996). *Modern society shocked by its risks, Vol. 17*, University of Hong Kong Department of Sociology Occasional Papers1–19.
Lyell, C. (2009). *Principles of geology, Vol. 1*. Oxford: Oxford University Press.
de Maillet, B. (1750). *Telliamed, or, conversations between an Indian philosopher and a French missionary on the diminution of the sea.* London.
Maupertuis, P.-L. (1742). *Lettre sur le comete.* Paris.
Mayor, A. (2000). *The first fossil hunters: Palaeontology in Greek and Roman Times.* Princeton: Princeton University Press.
Medwin, T. (1824). *Conversations of Lord Byron.* London.
Meijer, M. C. (1999). *Race and aesthetics in the anthropology of Petrus Camper (1722–1789)*. Amsterdam: Rodopi.
Merchant, C. (2016). *Autonomous nature: Problems of prediction and control from ancient times to the scientific revolution.* London: Routledge.
Futures literacy: Transforming the future. In R. Miller (Ed.). *Transforming the future: Anticipation in the 21st century* (pp. 1–12). London: Routledge.
Milne, D. (1828). *Essay on comets.* Edinburgh.
de Moivre, A. (1718). *The doctrine of chances; or, a method of calculating the probability of events in play.* London.
Montesquieu, C. (2008). *The Persian letters (M. Mauldon, trans.)*. Oxford: Oxford University Press.
Napier, W. (2008). Hazards from comets and asteroids. In N. Bostrom, & M. M. Ćirković (Eds.). *Global catastrophic risks* (pp. 222–237). Oxford: Oxford University Press.
Nares, E. (1834). *Man, as known to us theologically and geologically.* London.
O'Connor, R. (1991). Mammoths and maggots: Byron and the geology of cuvier. *Romanticism, 5*, 26–42.
Olbers, W. (1810). Über die Möglichkeit, daß ein Komet mit der Erde zusammenstoßen könne. *Monatliche Correspondenz zur Beförderung der Erd- und Himmelskunde, 22*, 409–450.
Parejko, K. (2003). Pliny the Elder's Silphium: First recorded species extinction. *Conservation Biology, 3*, 925–927.
Plato (2004). *Protagoras and Meno (R.C. Bartlett, trans.)*. Ithaca: Cornell University Press.
Pope, A. (2016). *Essay on man.* Princeton: Princeton University Press.
Ray, J. (1686). *Historia plantarum generalis.* London.
Rees, M. (2009). *Our final hour: A scientist's warning.* New York: Basic Books.
Rescher, N. (1998). *Predicting the future: An introduction to the theory of forecasting.* Albany: State University of New York Press.
Roberts, C. (2007). *The unnatural history of the sea.* Washington: Island Press.
Rolfe, W. D. (1985). William and John hunter: Breaking the great chain of being. In W. Bynum, & R. Porter (Eds.). *William hunter and the eighteenth-century medical world* (pp. 207–226). Cambridge: Cambridge University Press.
Rosendahl, W., Kempe, S., & Doppes, D. (2005). The scientific discovery of the *Ursus spelaeus*. *Naturhistorische Gesellschaft Nürnberg, 45*, 199–214.
Rudwick, M. J. S. (1975). Caricature as a source for the history of science: De la Beche's Anti-Lyellian Sketches of 1831. *Isis, 66*, 534–560.
Rudwick, M. J. S. (2008). *Georges Cuvier, fossil bones, and geological catastrophes: New translations and interpretations of the primary texts.* Chicago: University of Chicago Press.
de Sade, D. A. F. (1965). *Justine (A. Wainhouse, trans.)*. New York: Grove/Atlantic.
de Sade, D. A. F. (1968). *Juliette (A. Wainhouse, trans.)*. New York: Grove/Atlantic.
Sauer, M. (1802). *An account of a geographical & astronomical expedition to the northern parts of Russia.* London.
du Séjour, A. P. D. (1775). *Essai sur les comètes en general: et particulierement sur celles qui peuvent approcher de l'orbite de la terre.* Paris.
Shelley, M. W. (1999). *Frankenstein, or the Modern Prometheus, the 1818 version.* Ontario: Broadview Press.
Stafford, F. (1994). *The last of the race: The growth of a myth from Milton to Darwin.* Oxford: Oxford University Press.
Stewart, P. (1986). Science and superstition: Comets and the French public in the 18th century. *American Journal of Physics, 54*, 16–24. https://doi.org/10.1119/1.14763.
Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioural and Brain Sciences, 30*, 299–313. https://doi.org/10.1017/S0140525X07001975.
Swanson, L. R. (2016). The predictive processing paradigm has roots in Kant. *Frontiers in Systems Neuroscience, 10*. https://doi.org/10.3389/fnsys.2016.00079.

Torres, P. (2017). *Morality, foresight and human flourishing: An introduction to existential risks.* Durham: Pitchstone Publishing.
Tulving, E. (2002). Chronosthesia: Conscious awareness of subjective time. In D. T. Stuss, & R. T. Knight (Eds.). *Principles of frontal lobe function*. Oxford: Oxford University Press.
Turvey, S. T., & Cheke, A. S. (2008). Dead as a Dodo: The fortuitous rise to fame of an extinction icon. *Historical Biology, 20*, 149–163.
Voltaire, F. M. A. (1765). *The philosophical dictionary, from the French (anon., trans.).* Glasgow.
Wagar, W. W. (1982). *Terminal visions: The literature of last things.* Bloomington: Indiana University Press.
Wright, T. (1750). *An original theory or new hypothesis of the universe.* London.
Wright, T. (1968). *Second or singular thoughts upon the theory of the universe.* London: Dawsons of Pall Mall.
Zach, F. (1802). Forgesetzte Nachrichten uber den neuen Haupt-Planeten unseres Sonnen-Systems, Pallas Olbersiana. *Monatliche Correspondenz, 6*, 71–96.