

# Existential risks from a Thomist Christian perspective

Stefan Riedener

Global Priorities Institute | January 2021

*GPI Working Paper No. 1-2021*



# Existential risks from a Thomist Christian perspective

*Stefan Riedener, University of Zurich*

## 1 Introduction

Let's say with Nick Bostrom that an 'existential risk' (or 'x-risk') is a risk that 'threatens the premature extinction of Earth-originating intelligent life or the permanent and drastic destruction of its potential for desirable future development' (2013, 15). There are a number of such risks: nuclear wars, developments in biotechnology or artificial intelligence, climate change, pandemics, supervolcanos, asteroids, and so on (see e.g. Bostrom and Ćirković 2008). So the future might bring

**Extinction:** We die out this century.

In fact, Extinction may be more likely than most of us think. In an informal poll, risk experts reckoned that we'll die out this century with a 19% probability (Sandberg and Bostrom 2008). The Stern Review on the Economics of Climate Change, commissioned by the UK government, assumed a 9.5% likelihood of our dying out in the next 100 years (UK Treasury 2006). And a recent report by the Global Challenges Foundation suggests that climate change, nuclear war and artificial intelligence alone might ultimately result in extinction with a probability of between 5% and 10% (Pamlin and Armstrong 2015).<sup>1</sup> But the future needn't be so grim. It may also bring

**Survival:** We survive for another billion years, and on average there are always 10 billion people, who live very good 100-year-long lives. So there are 100 million billion future people with very good lives.

This may sound optimistic. But it's also possible. The earth seems to remain sustainably inhabitable by at least 10 billion people (United Nations 2001, 30), and for another 1.75 billion years (Rushby et al. 2013). The quality of our lives seems to have increased continuously (see e.g. the data collected in Pinker 2018), and it seems possible for this trend to continue. Notably, it partly depends on us

---

<sup>1</sup>More precisely, the report distinguishes between 'infinite impact', 'where civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends', and an 'infinite impact threshold', 'an impact that can trigger a chain of events that could result first in a civilisation collapse, and then later result in an infinite impact' (2015, 40). The above numbers are their estimates for infinite impact thresholds.

whether Extinction or Survival will materialise. In fact it may depend on what we do *today*. We could now promote academic research on x-risks (Bostrom and Ćirković 2008), global political measures for peace, sustainability or AI safety (Cave and ÓhÉigeartaigh 2019), the development of asteroid defence systems (Bucknam and Gold 2008), shelters (Hanson 2008), risk-proof food technologies (Denkenberger and Pearce 2015), and so on. And while none of these measures will bring x-risks down to zero, they'll arguably at least reduce them. So all of this raises a very real practical question. How important is it, morally speaking, that we now take measures to make Survival more likely?

The answer depends on the correct moral theory. It's most straightforward on standard utilitarianism. Suppose we increase the probability of Survival over Extinction by just 1 millionth of a percentage point. In terms of overall expected welfare (setting nonhuman sentience aside), this is equivalent to saving about 1 *billion* lives, with certainty. So according to utilitarianism, even such tiny increases in probability are astronomically important. Indeed, Nick Bostrom suggested that x-risk reduction 'has such high utility that standard utilitarians ought to focus *all* their efforts on it' (2003, 308ff.; see also Parfit 1984, 452f., Beckstead 2013). But this implication isn't restricted to utilitarianism. Something very similar will emerge on any view that assigns weight to expected impartial welfare increases. Consider Effective Altruism (or 'EA'). Effective Altruism is the project of using evidence and reasoning to determine how we can do the most good, and taking action on this basis (see MacAskill 2015). This doesn't presuppose any specific moral theory about what the 'good' is, or about what other reasons we have beyond doing the most good. But Effective Altruists typically give considerable weight to impartial expected welfare considerations. And as long as we do, the utilitarian rationale will loom large. Thus according to a 2018 survey, EA-leaders consider measures targeted at the far future (e.g. x-risk reduction) 33 times more effective than measures targeted at poverty reduction (Wiblin and Lempel 2018). The EA-organisation 80'000 hours suggests that 'if you want to help people in general, your key concern shouldn't be to help the present generation, but to ensure that the future goes well in the long-term' (Todd 2017). And many Effective Altruists already donate their money towards x-risk reduction rather than, say, short term animal welfare improvements.

In this paper, I'll ask how the importance of x-risk reduction should be assessed on a Christian moral theory. My main claim will be that Christian morality largely agrees with EA that x-risk reduction is extremely important—albeit for different reasons. So Christians should emphatically support the abovementioned measures to increase the probability of Survival.

Let me clarify. First, there's no such thing as *the* Christian moral doctrine. One of the philosophically most elaborate and historically most influential developments of Christian thought is the work of Aquinas. So I'll take this as my starting point, and argue first and foremost that core *Thomist* assumptions support x-risk reduction. Thomas's specific interpretation of these assumptions are often unappealing today. But I'll also claim that they can be interpreted more plausibly, that their core idea is still authoritative for many Christians, and that on any plausible interpretation they ground an argument for x-risk reduction. So while I'll start with Thomas, my conclusions should appeal to quite many contemporary Christians. Indeed, I'll ultimately indicate that these assumptions needn't even be interpreted in a specifically Christian manner, but emerge

in some form or other from a number of worldviews (cf. section 4). Second, there are different x-risk scenarios, and they raise different moral issues. I think the case is clearest for ways in which humanity might literally go extinct, before being superseded by non-human intelligence, and as a direct consequence of our own actions. I'll refer to these cases as 'non-transitional anthropogenic extinction', and it's on these cases that I'll focus. It would be interesting to explore other scenarios: cases in which we're replaced by another form of intelligence, non-extinction 'x-risks' (which Bostrom's definition includes) like an extended stage of suffering, or scenarios of natural extinction like volcano eruptions. My arguments will have upshots for such cases too. But I won't explore them here. Third, there are different ways in which, or different agents for whom, x-risk reduction might be 'important'. In what follows, I'll mostly be considering a collective perspective. That is, I'll assume that we as humanity collectively do certain things. And I'll focus on whether *we* ought to do anything to mitigate x-risks—rather than on whether you individually ought to. The existence of this collective perspective might be controversial. But I think it's plausible (see e.g. Dietz 2019). Christian moral theory, or at least Thomas, also seems to assume it (cf. section 2.1). And many important issues emerge only or more clearly from it. So I'll assume it in what follows.

In short, my question is about how important it is, on roughly Thomist premises, for us to reduce risks of non-transitional anthropogenic extinction. I'll first present three considerations to the effect that, if we *did* drive ourselves extinct, this would be morally very problematic—a hubristic failure in perfection with cosmologically bad effects (section 2). I'll then discuss some countervailing considerations, suggesting that even though such extinction would be bad, we needn't take measures against it—because God won't let it happen, or because we wouldn't *intend* it, or because at any rate it isn't imminent (section 3). I'll argue that none of these latter considerations cut any ice. So I'll conclude that on roughly Thomist premises it's extremely important for us to reduce x-risks.

## 2 Three Thomist Considerations

There are many Thomist considerations that would bear on x-risks. For instance, in driving ourselves extinct, we'd presumably kill the last representatives of humanity. On a Thomist perspective, those killings will be morally problematic simply *as killings* (see ST, ii-II, q64). Yet this has nothing to do with the fact that those killings lead to extinction. So let's see whether there would be anything *distinctly* problematic about extinction, if we brought it about. There is.

### 2.1 The natural law

A first consideration follows from Thomas's teleology, or from the Aristotelian strand in his thinking. Recall that for Thomas the order of the cosmos is teleological. This teleology is grounded in the fact that the cosmos is subject to God. And it takes the form of a law for us: 'the universe is governed by Divine Reason. Wherefore the very Idea of the government of things in God [...] has the nature of a law.' (ST, i-II, q91 a1, co) For Thomas, following this 'eternal law'—or Divine 'will' (ST, i-II, q93 a4, ad1) or 'plan' (ST, i-II, q93 a3,

co) for all things—is *the* ethical imperative for us. So what does it command? We can't know God's intent 'in itself' (ST, i-II, q93 a2, co), at least not in this earthly life (see ST, i-II, q93 a2, co; ST, I, q12 a11). But we can know it 'in its effect', through its manifestations in creation, or through the 'natural law' that we recognise as structuring the physical universe. In particular, we can detect God's plan for us through the natural inclinations He's implanted in us: 'all things partake somewhat of the eternal law[:] [...] from its being imprinted on them, they derive their respective inclinations to their proper acts and ends' (ST, i-II, q91 a2, co). In other words, our inclinations allow an indirect cognition of the essence of God's will, quite like sunrays allow an indirect cognition of the substance of the sun (ST, i-II, q93 a2, co). So what are our natural inclinations? Thomas speaks of three kinds: 'in man there is first of all an inclination [...] which he has in common with all substances: [...] the preservation of its own being [...]. Secondly, there is in man an inclination [...] which he has in common with other animals: [...] sexual intercourse, education of offspring and so forth. Thirdly, there is in man an inclination to good, according to the nature of his reason [...]: [...] to know the truth about God' (ST, i-II, q94 a2, co). What's crucial for present purposes is the second type of inclination. Procreation or the 'preservation of the species' is firmly part of our 'natural good' (SCG, III, 122-9), or of what the natural law commands us to do.

Now Thomas doesn't believe that everyone must follow all of these inclinations, or that everyone must have offspring. It's permissible for some to remain celibate (see e.g. ST ii-II, q152). But we as a collective have a duty—a 'duty [...] to be fulfilled by the multitude' (ST ii-II, q152 a2, ad1)—to procreate. So this grounds a straightforward consideration for x-risk reduction. By going extinct, we'd fail to attain our end, or to accord with the natural law. Indeed, our failure would be profound. It wouldn't just be some of us flouting this law—the bad apples in an overall virtuous whole. We'd fail *collectively*, as an entire species, to attain our end. And (at least as far as this life's concerned) we wouldn't just fail in one aspect of perfection, while still able to excel in others. Our survival is a precondition for any aspect of our flourishing. So our extinction would mean we fail *comprehensively*, in all respects of our end. And of course we wouldn't just temporarily fall short of our calling. Once we've gone extinct, there's no second bite at the apple. It would mean we've foundered *irreversibly*. So the moral failure in anthropogenic extinction would seem *complete*. In short:

**Natural law:** We ought to attain our natural end. Our extinction would prevent us from doing so—collectively, comprehensively, and irreversibly. Thus non-transitional anthropogenic extinction would amount to a total moral failure of us as a species.

That's a first consideration for why our extinction would be problematic.

Now I suppose that this kind of rationale isn't parochially *Thomist*, but should appeal to Christians quite broadly. Thomas himself interprets the natural law pertaining to procreation very radically—e.g. as permitting a non-procreative life only for the sake of the 'contemplation of truth' (ST ii-II, q152 a2, co; also ST ii-II, q153 a2, co),<sup>2</sup> and as prohibiting any intercourse known to

<sup>2</sup>Thomas doesn't say this explicitly in these passages. He says that the goods of the body are subordinate or 'directed' to the goods of the soul, and that virginity (or a lack of bodily good) is justified if done for the sake of Divine contemplation (or good of the soul; ST ii-II, q152 a2, co). I'm reading the conditional here as a biconditional.

be non-reproductive (SCG, III, 122; also ST ii-II, q153 a2, co). This would mean that our procreation-related obligations go much beyond preventing extinction. And it would mean that contemporary liberal moral thought is radically wrong about the good human life, and about intercourse among people of the same sex, or people who for whatever reasons cannot or don't want to reproduce. Many contemporary Christians will want to resist these implications. But we needn't understand the general idea of a 'natural law' in this manner. Plausibly, other aspects of the human end (athletic, social, emotional) are just as integral as 'the contemplation of truth'. So there are good lives without children beyond those of philosophers and priests (see e.g. Nussbaum 1987 or 2011). Also, perhaps there are other functions of human sexuality (e.g. bonding), such that non-reproductive intercourse isn't a misuse of sexual organs. After all, there's plenty of non-reproductive sexuality among non-human animals (see e.g. Bagemihl 1999). The details of the natural law are a matter of large debate.<sup>3</sup> But I presume that the *general* idea of a 'natural law' is still quite prominent for Christians today. Indeed, assuming the universe manifests God's intentions, it's a very natural assumption. And I suppose that on any plausible interpretation, it will ground a consideration against extinction: whatever it implies about childfree individuals, contraception or homosexuality, it seems hard to square the natural law with an heirless self-eradication of our species. If the law commands anything, it seems, it commands us to 'be fruitful and multiply' (Gen 1:28). So a consideration along these lines should be authoritative to Christians quite broadly. Self-inflicted extinction would constitute a total failure in our executing our designated role.<sup>4</sup>

## 2.2 Humility

However, there's more to anthropogenic extinction than a failure to reproduce. Let's look at the precise way in which, through anthropogenic extinction, we'd fail to attain our end. According to Thomas, there are different ways to fall short of perfection. One is through utter passivity, or sloth, or a 'sorrow' that 'so oppresses man as to draw him away entirely from good deeds' (ST, ii-II, q35 a2, co). So we might just not bother to do anything much at all, and therefore fail in perfection. Another way of failing lies in falling too low, or being overcome with 'the lower appetite, namely the concupiscible' (ST, ii-II, q153 a5, co). So we might behave like lower animals, and fail to live *up* to our standards. But drawing on Augustine, Thomas says that 'the most grievous of sins' (ST, ii-II, q162 a6, co) consists in aiming too *high*—in failing to respect our limitations, or acting as if we were God. To do so is to commit the sin of pride, hubris, or *superbia*. Following the church father (De Civ. Dei XIV, 13), Thomas characterises a prideful man as someone who 'aims higher than he is', or does not 'tend to that which is proportionate to him' (ST, ii-II, q162 a1, co). And what's proportionate to us is of course not so by coincidence, but due to Divine assignment. So pride is opposed specifically to humility:

---

<sup>3</sup>For a classic systematic exploration of natural law, see Finnis (2011). For an in-depth treatment of Aquinas's theory, see e.g.—among very many others—Lisska (1998).

<sup>4</sup>Historically, this broad rationale hasn't appealed to all Christians, of course. Paul himself seems to suggest it would be best if everyone was celibate (see 1 Cor 7:7). And the Cathars even thought reproduction was a moral evil. I thank Felix Timmermann and Christoph Halbig for pointing this out to me.

‘humility properly regards the subjection of man to God [...]. Hence pride properly regards lack of this subjection, in so far as a man raises himself above that which is appointed to him according to the Divine rule’ (ST, ii-II, q162 a5, co). And in this sense, as far as the ‘aversion from the immutable good’ (ST, ii-II, q162 a6, co) is concerned, pride is the most grievous sin: it’s not just a failure through ignorance or weakness or an innocent desire for another good, but an active ‘withstanding’ or ‘resisting’ or manifesting ‘contempt of’ God (ST, ii-II, q162 a6, co).

What does this imply in practice? Thomas specifies what he means. Pride isn’t any old undue desire. It’s, specifically, an ‘appetite for *excellence* in excess of right reason’—or an inordinate imitation of the powers or competences of God (ST, ii-II, q162 a1, ad2; emphasis added). This may take various forms. We may be pridefully curious about things we can’t know, such as facts about good and evil (ST, ii-II, q163 a1, ad3). Or we may unduly discard our need for Divine grace, deeming ourselves capable of happiness on our own (ST, ii-II, q163 a2, co). But a more specific power that isn’t appointed to us is decisions about life and death: ‘it belongs to God alone to pronounce sentence of death and life’ (ST, ii-II, q64 a5, co). Thus to kill someone, or (I take it) actively prevent them from coming into existence, is generally to show an appetite for a power that doesn’t pertain to us. And if all of this is true, then non-transitional anthropogenic extinction in particular would manifest an enormous form of pride. Note that on most such scenarios, we’d fall prey to technologies we were unable to control—nuclear weapons, artificial intelligence, biotechnology, or whatever. So our extinction would mean that we’d overestimated our mastery over these fabrications, and the invulnerability we could leverage in the face of them. It would mean we were prideful in the general sense of desiring an undue excellence. And the upshot of this would be, specifically, a life-death decision on an astronomical scale: preventing lives perhaps by the million billions. It would mean we were prideful in this more specific sense too. Or in short:

**Humility:** Non-transitional anthropogenic extinction would mean we overstrained our power as a species. And the upshot of it would be a life-death decision on an astronomical scale. Thus it would amount to an enormous form of *superbia*.

That’s a second consideration for why our extinction would be problematic.

And here too, I suppose this consideration should appeal to Christians quite broadly. Again, Thomas’s specific interpretation of God’s authority over life/death-decisions should be controversial. He suggests that, while permitting us to execute death penalty (ST, ii-II, q64 a2, co), this authority absolutely prohibits suicide (ST, ii-II, q64 a5, co), and prohibits abortion from the moment of ensoulment (see e.g. ST, I, q118f.)—which some people have claimed, according to Thomas’s metaphysical principles and the known facts about embryology, takes place at the moment of fertilisation (Haldane and Lee 2003, 273).<sup>5</sup> This would mean self-extinction can under no circumstances be permissible. And it would mean that contemporary liberal moral thought is radically wrong about death penalty, abortion, or suicide and euthanasia. Many contemporary Christians will want to resist these implications. But again we needn’t interpret the

---

<sup>5</sup>In the *Scriptum super Sententiis*, following Aristotle, Thomas himself suggests that the soul is infused after 40 days for males and after 90 days for females (SSS III, d3 q5 a2, co).

general idea of *superbia* in this manner. After all, it's implausible that criminals are 'dangerous and infectious to the community' and must be cut away like an infected part of a body (ST, ii-II, q64 a2, co). Perhaps suicide can in some instances of extreme pain, or loss of autonomy or dignity, be a manifestation of self-love. And if so, perhaps we can then view ourselves as authorised by God—or His commandment to love ourselves (e.g. Mat 22:39)—to end our lives. After all, Thomas himself (defending Abraham's intention to kill Isaac) suggests that 'he who at God's command kills an innocent man does not sin' (ST, ii-II, q64 a6, ad1). Perhaps some forms of abortion (e.g. after rape) can also be seen as an expression of self-love or -respect. Or perhaps we must ascribe to Aquinas a different view of ensoulment (Pasnau 2002), or simply reject some of his metaphysics in light of more recent discoveries. Again, the details of *superbia* are contested.<sup>6</sup> But I assume that the pertinent general idea is still prominent among Christians today. Indeed, that playing God is a sin is a natural corollary of theism. The in-principle-ban on life/death-decisions has a very firm grounding in the Bible (Thomas cites Deut 32:39, for instance). And I suppose that on any plausible interpretation, these ideas ground a consideration against extinction: whatever they imply about death penalty, abortion, and suicide, the dictates of humility seem hardly compatible with our developing a technology that accidentally seals the fate of our whole species. So again, such a consideration should have import for Christians quite broadly. Self-inflicted extinction would constitute a shattering form of *superbia*.

### 2.3 The value of humanity

But we need to see another aspect of Thomas's view of humanity too, which is perhaps most distinctly Biblical, and which is implicit in the third kind of inclination he ascribes us. For sure, we're not God. But we are nonetheless made to know the truth about Him. In fact, of all corporeal things (i.e., disregarding angels) we're the only ones whose nature enables such knowledge. And this, for Thomas, marks our 'dignity'. It's this dignity which made it fitting for Jesus to adopt *human* nature, rather than becoming an animal or any other thing: 'in the irrational creature the fitness of dignity is wanting.' (ST, III, q4 a1, co) Indeed, it's these intellectual capacities that ground our likeness to God—a likeness greater than that of any non-reasoned creature (see ST, I, q93 a2, co), and sufficient to say we're His 'image': 'intellectual creatures alone [...] are made to God's image.' (ST, I, q93 a2, co) And it's this which ultimately manifests that God loves us more than any other thing. So Aquinas quotes Augustine approvingly: "God loves all things that He has made, and amongst them rational creatures more, and of these especially those who are members of His only-begotten Son Himself." (ST, I, q20 a3, sc)

In other words, and as Thomas says very explicitly (ST, I, q20 a4, co), of all

---

<sup>6</sup>There's a lot of recent literature concerning Thomas on humility. For longer treatments, see e.g. Tadie (2006) or Fullam (2001). As far as I see, Thomas himself doesn't discuss whether birth control or contraception infringes on God's dominion over life/death-decisions. But the Catholic church has since leveraged this argument. Yet perhaps a ban on contraception needn't follow from humility either. Perhaps Thomists may appeal to the doctrine of double effect, and compare well-intentioned prevention of fertilisation with well-intentioned killing in self-defense (cf. section 3.2). Or perhaps they might again claim to be authorised in this decision by the commandment to love ourselves, and love our partners, and our (potential) children for whom we couldn't be sufficiently responsible parents.



corporeal things we are the most valuable, or those with most ‘goodness’. And this difference is categorical. So our extinction would have cosmological ramifications. For standard utilitarians, there’s no categorical difference between a world populated by flourishing societies of people and a world populated by one lonely lizard basking in the sun and feeling a tinge of pleasure. The difference is a matter of degree. For Aquinas that’s different. In a world devoid of higher intelligence, there’s nothing that’s made to God’s image—nothing with our special dignity or perfection. This would radically undercut the perfection of the universe, as God created it, which depends on the varieties of goodness (see e.g. ST, I, q47). So our extinction would change the face of creation. Or in short:

**The value of humanity:** Non-transitional anthropogenic extinction would constitute the destruction of the most valuable aspect of creation—and thus a loss of categorical cosmological significance.

That’s a third consideration for why our extinction would be problematic.

Again, I suppose that this consideration should appeal to Christians very broadly. Thomas interprets the order of creation very radically. He suggests that everything else exists *for us* (ST, ii-II, q64 a1, co)—such that we may use animals simply as means for our ends, say, and the self-interested killing of another person’s ox wrongs at most its owner (ST, ii-II, q64 a1, co). This would mean that our extinction would literally destroy the purpose of the physical universe. And it would mean that much contemporary thought is radically wrong about animals, or other aspects of the natural world. Again, many Christians will want to resist these implications, or so I hope. But the distinct value of humans needn’t be interpreted this radically. Perhaps our ‘dominion’ over animals (Gen 1:26) doesn’t mean we can use them simply as means. Perhaps it means we must use our *de facto* power in the manner of a loving and respectful and liberal guard (see e.g. Linzey 2016, ch. 2). And similarly with every other aspect of creation—ecosystems, plants, and the climate. Again, the exact nature of our status is a large question. But the general idea of our special value still seems very prominent. It does seem hard to avoid if we take seriously that man is ‘God’s image’, or has a special (perhaps responsibility-implying) dominion over the earth—and very plausible given God’s human embodiment in Christ. And on any interpretation, our special status will ground a consideration against extinction: whatever it implies about our responsibilities towards animals, the face of creation will be changed categorically if we drive ourselves extinct. So this consideration should be authoritative to Christians very broadly. Self-inflicted extinction would constitute a destruction with cosmological import.

Plausibly, there are other Christian considerations that bear on x-risks, at least once we move beyond Thomas. Most notably, perhaps Christian *caritas* simply gives us reasons to increase expected impartial welfare (see e.g. William Paley’s utilitarian Christianity, in his *Principles of Moral and Political Philosophy*; Paley 2014, esp. book II). If it does, the EA rationale will be very pertinent to Christians as it stands. But whether that’s an apt interpretation of the Christian virtue—or whether, as I personally find more likely, the latter is more about making people happy than about making happy people—will be controversial. And at any rate, I think these are the three most distinctly Christian concerns. So let me leave it at that for now.

### 3 Discussion

If all of this is right, non-transitional anthropogenic extinction would be a morally highly problematic result. Such extinction wouldn't just amount to a regrettable form of imprudence—but to a failure to fulfill our God-given role. It wouldn't amount to any old failure of this kind—but to a prideful miss of our end. And it wouldn't be just an inconsequential overstepping of our dominion—but an unauthorised decision with categorical cosmological ramifications.<sup>7</sup> But as mentioned, humanity can now (in the form of the present generation) take measures to reduce the likelihood that it will eventually effect this result. So this suggests that we have very strong reasons to take some such measures. It suggests that Christians too have strong reasons to donate their money towards x-risk reduction rather than, say, disaster relief; to conduct x-risk research rather than enquiries about the cause of the grief of a neighbour; or to advocate political measures for long-term safety rather than for short-term caritative purposes. But let's not get ahead of ourselves. Some considerations in Thomas's philosophy seem to suggest that even if extinction would, as far as we know, be bad, it's not important that we now do much about it in practice. So let me turn to some of these theological and moral countervailing concerns. This will not simply corroborate the results so far. It will also help clarify the precise form and weight of what the above rationale implies.

#### 3.1 Theological considerations: Divine providence

One apparent reason against x-risk activism is Divine providence. Again, for Thomas, everything is subject to God. And Thomas's God is all-knowing, all-powerful, and all-good: 'In God there exists the most perfect knowledge' (ST, I, q14 a1, co), 'He can do all things that are possible absolutely [or don't imply a contradiction in terms]' (ST, I, q25 a3, co); and He 'loves all existing things' (ST, I, q20 a2, co). One might conclude from this that God won't allow us to go extinct, or that if He does it's somehow just for the better. And from this one might infer that we needn't do anything to prevent our extinction in practice. X-risk reduction is obviated by the rule of a loving God.

Now this is difficult and well-trodden terrain. But God's existence doesn't *in general* seem to warrant any heedlessness. Consider risks of road accidents, say. And suppose any accident accords with His will. It's an interesting question what precisely this implies. Perhaps it warrants a certain ultimate serenity or comfort: a reassurance at the thought that whatever happens on our roads has its mysterious rightness in the grander scheme of things. Perhaps it warrants a fundamental form of trust: a consolation in the belief that our diligence will be duly rewarded (see ST, ii-II, q22 a1, co). But God's unfathomable values cannot act as a guide for us, or as an excuse to depart from the norms and expectations we're given. So these forms of Christian hope must be distinguished from optimism that accidents won't occur, or from feeling exonerated from the need to take caution. Christians ought to see to it that they stop at red lights, respect speed limits, and check their brakes—and just as carefully as everyone else. Indeed, Thomas is the first to emphasise that we're obligated to such

---

<sup>7</sup>Perhaps it's worth noting that in principle, these considerations are logically independent. One might accept some but not all of them. For instance, one might doubt that our natural inclinations reveal a Divine law, but still accept the cosmological disvalue of human extinction.

circumspection (see e.g. ST, ii-II, q64 a8, co). In short, Divine providence doesn't affect our need to reduce road risks in practice. But then it's unclear why it should do so with risks of human extinction.<sup>8</sup>

One might suggest that there's something special about human extinction, setting it apart from everyday hazards, and making it particularly unlikely. Perhaps it's precisely because we're His 'image', or uniquely most valuable, that God won't allow us to perish. Perhaps Jesus's redemption of mankind would have been futile if some sorrowful two thousand years later He let us burn up or wither away.<sup>9</sup> Or perhaps, more specifically, there's evidence in scripture that we'll survive: the primal great flood precisely *didn't* erase us, and after that deluge there seems to have been a promise of survival—that 'never again will all life be destroyed by the waters of a flood.' (Gen 9:11) This isn't the place to dive deep into Christian theology. But it's hard to see why such reasonings should give us much confidence. The Lord moves in mysterious ways. If all known misery is compatible with His providence, it must surely at least be possible that He'd allow us go extinct. In particular, there are many more passages in scripture painting a grim prospect of extinction than promising a boundless glorious future. And many such visions of the 'end of time' indicate anthropogenic extermination—that 'the nations in the four corners of the earth' are gathered for 'battle' or 'war' (Rev 20:8), right before 'the earth and the heavens' are gone (Rev 20:11). If anything, according to the Bible, anthropogenic extinction seems a very live option. So Divine providence might have implications for the metaphysics and ultimate unfathomable significance of our doomsday. But it doesn't seem to warrant *our* heedlessness about x-risks in practice. If the above considerations are sound, we, just as any generation in the future, must see to it that it isn't *us* who bring a Johannine finale about.

### 3.2 Moral considerations: deontology

Let's look at moral considerations that might mitigate the importance of x-risk reduction. One such consideration concerns intentions. For Thomas, the permissibility of our actions can depend on what we intend. More precisely, he held, or indeed introduced, the doctrine of double effect (or 'DDE'). According to this doctrine, there's a difference between the effects we intend an action to have, and the effects we merely foresee but don't specifically intend. While it's always impermissible to intend evil effects, it can be permissible to cause them, if what you intend is good. For instance, you may kill in self-defence, if you intend to save your life and merely foresee the death of your aggressor (see ST, ii-II, q64 a7, co). This bears on x-risks. Perhaps non-transitionary anthropogenic extinction would be a very bad result. But one might suggest that as long as we don't *intend* to effect it, we wouldn't have acted impermissibly: an unintentional

---

<sup>8</sup>In his contribution to this collection, Dominic Roser suggests that Christianity warrants a certain renunciation of control. I'm not sure whether he intends this to contradict my claims here—and to imply, say, that Christians may reasonably check their brakes half as often as atheists. If so, his view seems to me, among other things, unstable. It seems that God's existence would either require standard full-blown caution (as I think Thomas suggests), or would warrant *total* relinquishment of it. It seems hard to see why He would warrant a certain but *only* a certain trust in the course of events. At any rate, I'd think most Christians don't understand their faith in this manner. And even if Roser is right, any such *limited* trust in things would arguably still not undermine my practical conclusions.

<sup>9</sup>I thank Felix Koch for mentioning this thought to me.

mass life-prevention isn't prideful, and an unintentional thwarting of perfection or erasing of God's image isn't morally wrong. So if our intentions are fine, we needn't worry about x-risk reduction.<sup>10</sup>

But this argument is a non-starter. It might well be worse to intend our extinction rather than to merely foresee it. For Thomas, it would mean that our action would lack 'goodness from its end', and not just from its circumstances or species (ST, i-II, q18, esp. a7, co). But this doesn't mean that unintentional extinction wouldn't be wrong, or even specifically prideful. There's no blanket permission for the well-intentioned. You mustn't speed through the cross-walk and put people at risk, even with the laudable aim of being punctual at the faculty meeting. Thomas says explicitly that you are 'in a sense guilty of voluntary homicide'—and thus (I take it) of *superbia*—if you kill someone unintentionally but without having taken 'due care' (ST, ii-II, q64 a8, co). He doesn't specify when precisely the DDE allows you to accept a foreseen evil. But he suggests it depends on (*i*) the expected goodness of the intended good, (*ii*) the expected badness of the foreseen evil—or on the 'proportion' between the two (ST, ii-II, q64 a7, co)—and (*iii*) on whether the harmful action is 'necessary' (ST, ii-II, q64 a7, co) for the good, or whether there are ways of securing the good without these bad effects. These criteria decidedly command x-risk reduction. Sure, many technologies that involve such risks also promise important goods. But if my arguments are right, the moral costs are potentially tremendous. And there are less risky courses of action which detract little from the expected good. The above-mentioned measures would reduce x-risks, but wouldn't really jeopardise the benefits these technologies promise. So it seems absolutely obligatory to take them. Going on as we do does seem like scorching through the cross-walk for timeliness at the meeting. That isn't killing for the sake of the killing. But it's a serious violation of the negative constraint of the DDE. And thus it's surely grave enough.<sup>11</sup>

Let's consider another countervailing consideration. Perhaps extinction would be terrible. But for all we know, it isn't actually imminent. Indeed, compared to more immediate moral callings like global poverty, gender justice, or disaster relief, it seems very remote. According to utilitarianism or EA, such distance doesn't matter. These views are absolutely impartial. But perhaps Christian morality is different. One might suggest Christian morality is more concerned with visible, immediate, near-at-hand moral problems—the wounded person alongside the road (Luke 10:30), or perhaps the literal 'neighbor' (Mark 12:31)—and doesn't warrant too much concern about such far-distant matters. In fact, Thomas himself is explicit that we should be partial, at least in the virtue of *caritas*: 'Among our neighbors, we should love them more who are

---

<sup>10</sup>I thank Eric Sampson and Jonathan Erhardt for mentioning this thought to me.

<sup>11</sup>Strictly speaking, perhaps extinction wouldn't even be a 'foreseen' effect of our actions. By 'foreseen' [*praecogitatus*] Thomas seems to mean 'known with certainty to result' (see e.g. ST, i-II, q20 a5, co). But we aren't certain that promoting AI, say, will result in extinction. We're just not certain that it won't. In discussing the DDE, Thomas doesn't deal with uncertainty. But he does elsewhere. In ST, i-II, q20 a5, co, he says: 'if the consequences [of an action] are not foreseen, we must make a distinction. Because if they follow from the nature of the action and in the majority of cases [...] the consequences increase the goodness or malice of that action [...]. On the other hand, if the consequences follow by accident and seldom, then they do not increase the goodness or malice of the action.' This can't be right. Throwing stones off a cliff might kill passersby below in only 5% of cases. But then it's nonetheless wrong. Thomas lacked the concept of expected value. Today, we'd surely interpret the doctrine in terms of that concept, or some related one.

more closely connected to us' (DQV, q2 a9, co; cf. ST ii-II, q31 a3, co). So one might conclude that although our extinction would be bad, we needn't worry about x-risk reduction—or at least not now, or not with resources we could direct towards these more immediate concerns.

But this would be a mistake. Christian morality may be partialist about *caritas*. So there may be a tension between Christian benevolence and the impartialist EA-directive to 'do the most good'. But on the considerations I've sketched, reducing x-risks isn't a form of 'doing good', or of charity, or of fulfilling any positive obligation. It's to ensure—or make more likely—that we don't flout our end, overstep our dominion, or wreck the crown of creation. Formally, it's to ensure we satisfy the constraint of the DDE. In other words, it's to respect a perfect negative duty. And Thomas doesn't seem to endorse partiality or discounting in such negative duties, and neither does Christian morality more broadly. On Christian morality, killing, say, is wrong. And doing what will kill someone in a month is presumably *ceteris paribus* just as wrong as doing what will kill someone in a year, or in a thousand years for that matter. So the sheer temporal distance of human extinction *in itself* doesn't seem to obviate x-risk reduction. On the contrary, note that the negative character of this obligation has implications for how it may be weighed against others. For utilitarians, the obligation to reduce x-risk is formally on a par with (other) obligations of benevolence, such as obligations to support the poor. So it ought to be weighed against them. For Christians, if I'm right, this is different. As a negative obligation, the obligation to reduce x-risks must not be weighed against positive imperfect obligations, or can't be discharged by doing enough by way of benevolence elsewhere. It's quite simply a perfect duty.

There's a final point worth noting. For utilitarians, notoriously, we may take any means for the sake of the good end. If killing an innocent scientist reduces x-risks, we presumably ought to kill them. For Thomist Christians this will be different. Thomas explicitly accepts deontological constraints: 'some [actions] are evil, whatever their result may be.' (ST ii-II, q88 a2, ad2) So Thomist Christians must not do *anything* to reduce such risks. They generally mustn't kill or lie or steal for that purpose. In practice, however, this won't make much of a difference. The most salient means of x-risk reduction include academic research, or global measures for peace, sustainability or AI safety. None of these measures seem to violate any constraints. So even if Christian morality is deontological in nature, there's plenty of good ways to start making Survival more likely.

## 4 Conclusion

If all of this is right, it's not just that non-transitional anthropogenic extinction would, as far as we know, be a disastrous result—a cosmologically salient prideful miss of our end. At least in practice, and as a matter of a negative constraint against lack of 'due care' that in principle seems neither discountable nor weighable against positive obligations, Thomist Christians have very strong reasons to take deontologically permissible means to prevent such extinction: to conduct research on risks and risk-reduction, promote political arrangements with an eye to the very far future, donate heir money towards x-risk mitigation, and so on. Indeed, given the gravity of the possible effect, and the perfect or

*negative* form of the duty, they seem to have reason to do this even if it will considerably constrain their resources for classical forms of *caritas*.

Or that’s as far as our argument takes us. There are many considerations relevant to non-transitional anthropogenic extinction that I haven’t yet addressed. For instance, it would be interesting to explore the implications of Christian love for our question; to consider the relevance of other Christian virtues—such as temperance, diligence, or patience; or to integrate the assumption that in some form or other we’ll always live on, or that in this sense we can’t really die out.<sup>12</sup> That’s beyond the scope of this paper. It would also be interesting to consider the implications of my arguments for issues beyond my core question. Take risks of natural extinction. If we ought to ensure the preservation of the species, we must generally guard ourselves also against asteroid hits. It would presumably be a form of *superbia* to presume that no such event could erase us. And to let it happen would mean to let creation’s most valuable part be destroyed. Allowing natural extinction might not be as grave as actively extinguishing ourselves, but still seems a large-scale moral failure. Another, and more intricate question, concerns the issue of non-human higher intelligence. Suppose we turned into a non-human AI and left mankind behind. Would we then miss our end (in the non-preservation of the species), or would we precisely fulfill it (in the perfection of our ingenuity)? Would it be a form of *superbia* to thus intervene in the nature of species, or would that belong to our natural proportionate capacities? And would the most valuable aspect of the universe be lost, or would the cosmos become more valuable if inhabited by a more perfect form of reason?<sup>13</sup> Again, these are questions for another occasion.

There’s also the question whether similar arguments emerge on other religious or non-religious worldviews. They arguably do. The normative core of the first two arguments is simply a kind of perfectionism; and the core of the third argument is the idea of a special value of humanity. In Christianity these assumptions are grounded in the existence of the Christian God. But they needn’t be. In some form or other, you can adopt them without believing in that God, or indeed in any form of theism. We might see them as emerging, at bottom, from very general aspects of the Christian view of ourselves: from the ambivalent sense that there’s something greater than us, but that we nonetheless are quite great indeed. Perfectionism expresses the idea that what structures the universe exceeds mankind and is somehow commanding for us—so that we ought to fulfill our role in the scheme of things, and accept the limitations of it. The idea of our special value expresses that we have a certain greatness nonetheless—that our role isn’t any old role, and that it would *matter* if we were gone. It’s at bottom this dichotomous nature, fallen and fragile yet valuable, which implies we should be exceedingly careful with ourselves. Christians, and everyone with a similar view of mankind, and of our obligations and limitations and value, should agree with Effective Altruists that x-risk reduction is extremely important.

**Acknowledgments.** For helpful comments on earlier versions of this paper, I thank the participants of the workshop ‘Religious Perspectives on Effective Altruism’, the participants of the 5th Oxford Workshop on Global Priorities Research, as well as Christoph Halbig, Samuel Hughes, Sebastian Muders, Dominic Roser, Philipp Schwind and Felix Timmermann.

---

<sup>12</sup>I thank Dominic Roser for mentioning this thought to me.

<sup>13</sup>I thank Mara-Daria Cojocaru and Carin Ism for mentioning these questions to me. For a generally positive Christian stance on enhancement, see e.g. Keenan (2014).

## References

- Aquinas, Thomas. 2010. *Disputed Questions on Virtue* [DQV]. Translated by Jeffrey Hause and Claudia Eisen Murphy. Indianapolis: Hackett.
- Aquinas, Thomas. 1920. *Summa Theologica* [ST], second and revised edition. Translated by Fathers of the English Dominican Province. London: Burns Oates & Washbourne.
- Aquinas, Thomas. 1955-57. *Summa Contra Gentiles* [SCG]. Translated by Anton C. Pegis, James F. Anderson, Vernon J. Bourke, Charles J. O’Neil, edited by Joseph Kenny O.P. New York: Hanover House.
- Aquinas, Thomas. *Scriptum super Sententiis* [SSS]. Accessed September 17, 2019. <http://www.corpusthomicum.org/iopera.html>.
- Augustine. 2017. *The City of God Books XIII and XIV* [De Civ. Dei]. Edited and Translated by Peter G. Walsh. Liverpool: Liverpool University Press.
- Bagemihl, Bruce. 1999. *Biological Exuberance: Animal Homosexuality and Natural Diversity*. New York: St Martin’s Press.
- Bostrom, Nick, and Milan M. and Ćirković, eds. 2008. *Global catastrophic risks*. Oxford: Oxford University Press.
- Bostrom, Nick. 2003. “Astronomical Waste: The opportunity cost of delayed technological development.” *Utilitas* 15(3): 308-314.
- Bostrom, Nick. 2013. “Existential Risk Prevention as Global Priority.” *Global Policy* 4 (1): 15-31.
- Beckstead, Nick. 2013. “On the overwhelming Importance of Shaping the Far Future.” PhD diss., Rutgers.
- Bucknam, Mark, and Robert Gold. 2008. “Asteroid Threat? The Problem of Planetary Defence.” *Survival* 50 (5): 141-156.
- Cave, Stephen, and Seán S. ÓhÉigeartaigh. 2019. “Bridging near- and long-term concerns about AI.” *Nature Machine Intelligence* 1: 5-6.
- Denkenberger, David, and Joshua M. Pearce. 2015. *Feeding Everyone No Matter What: Managing Food Security After Global Catastrophe*. London: Elsevier.
- Dietz, Alexander. 2019. “Effective Altruism and Collective Obligations.” *Utilitas* 31 (1):106-115.
- Finnis, John. 2011. *Natural Law and Natural Rights*. Oxford: Oxford University Press.
- Fullam, Lisa. 2001. *The Virtue of Humility: A Reconstruction Based in Thomas Aquinas*. Ph.D. Dissertation, Harvard University.
- Haldane, John, and Patrick Lee. 2003. “Aquinas on human ensoulment, abortion and the value of life.” *Philosophy* 78 (2):255-278.

- Hanson, Robin. 2008. "Catastrophe, social collapse, and human extinction." In *Global catastrophic risks*, edited by Nick Bostrom and Milan M. and Ćirković, 363-378. Oxford: Oxford University Press.
- Keenan, James F. 2014. "Roman Catholic Christianity, Embodiment and Relationality: Roman Catholic Concerns about Transhumanist Proposals." in: *Transhumanism and the Body: The World Religions Speak*, edited by Calvin Mercer and Derek F. Maher, 155-171. New York: Palgrave.
- Linzey, Andrew. 2016. *Christianity and the Rights of Animals*. Eugene: Wipf and Stock Publishers.
- Lisska, Anthony J. 1998. *Aquinas's Theory of Natural Law: An Analytic Reconstruction*. Oxford: Oxford University Press.
- MacAskill, William. 2015. *Doing good better: how effective altruism can help you make a difference*. New York: Gotham Books.
- Nussbaum, Martha C. 2011. *Creating Capabilities: The Human Development Approach*. Cambridge, MA: Belknap /Harvard University Press.
- Nussbaum, Martha C. 1987. "Non-Relative Virtues: An Aristotelian Approach." *Midwest Studies in Philosophy* 13 (1):32-53.
- Paley, William. 2014. *The Principles of Moral and Political Philosophy*. Cambridge: Cambridge University Press.
- Pamlin, Dennis, and Stuart Armstrong. 2015. "Global Challenges: 12 Risks that threaten human civilisation." Accessed August 8, 2019. <https://api.globalchallenges.org/static/wp-content/uploads/12-Risks-with-infinite-impact.pdf>.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Pasnau, Robert. 2002, *Thomas Aquinas and Human Nature*. Cambridge: Cambridge University Press.
- Pinker, Steven. 2018. *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*. New York: Penguin.
- Richter, Daniel, Rainer Grün, Renaud Joannes-Boyau, Teresa E. Steele, Fethi Amani, Mathieu Rué, Paul Fernandes, Jean-Paul Raynal, Denis Geraads, Abdelouahed Ben-Ncer, Jean-Jacques Hublin, and Shannon P. McPherron. 2017. "The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age." *Nature* 546: 293-296.
- Rushby, Andrew J., Mark W. Claire, Hugh Osborn, and Andrew J. Watson. 2013. "Habitable Zone Lifetimes of Exoplanets around Main Sequence Stars." *Astrobiology* 13 (9): 833-849.
- Sandberg, Anders, and Nick Bostrom. 2008. "Global catastrophic risks survey." *Future of Humanity Institute Technical Report*. Accessed May 15, 2019. <http://www.global-catastrophic-risks.com/docs/2008-1.pdf>.
- Tadie, Joseph Lawrence. 2006. *Between humilities: A retrieval of Saint Thomas Aquinas on the virtue of humility*. Ph.D. Dissertation, Boston College.



- Todd, Benjamin. 2017. "Presenting the long-term value thesis." Accessed May 15, 2019. <https://80000hours.org/articles/future-generations/>.
- UK Treasury. 2006. "Stern review on the economics of climate change." Accessed May 15, 2019. [https://webarchive.nationalarchives.gov.uk/20100407172811/http://www.hm-treasury.gov.uk/stern\\_review\\_report.htm](https://webarchive.nationalarchives.gov.uk/20100407172811/http://www.hm-treasury.gov.uk/stern_review_report.htm).
- United Nations, Department of Economic and Social Affairs, Population Division. 2001. "World population monitoring, 2001: population, environment and development." Accessed May 15, 2019. <https://www.un.org/en/development/desa/population/publications/pdf/environment/population-monitoring.pdf>.
- Wiblin, Robert, and Howie Lempel. 2018. "What skills or experience are most needed within professional effective altruism in 2018? And which problems are most effective to work on? New survey of organisational leaders." Accessed May 15, 2019. <https://80000hours.org/2018/10/2018-talent-gaps-survey/>.