Can Anti-Natalists Oppose Human Extinction?

The Harm-Benefit Asymmetry, Person-Uploading, and Human Enhancement

Abstract: This article outlines a novel philosophical position according to which people can (a) value the continued survival of humanity, and (b) oppose procreation on moral grounds. While these two propositions may appear contradictory, they need not be: life-extension technologies could enable members of a "final" human generation to live indefinitely long lives and, therefore, to avoid biological extinction. I call this position *no-extinction anti-natalism*. After exploring a range of arguments for (a) and (b), I turn to various challenges associated with attaining "functional immortality." These include whether procreation can be morally justified until life-extension technologies become available, as well as whether personal identity issues associated with attaining functional immortality problematize the anti-natalist component of my position. I conclude that this view ought to be taken seriously by those who believe that procreation is immoral.

1. Introduction

This paper argues that one can accept without contradiction the following three propositions: (i) it would be better if no people had ever existed, (ii) it would be better if there are no more people, and (iii) human extinction would constitute a terrible tragedy that we should strive hard to avoid. The argument that I present goes as follows:

(1) Coming into existence is always a net harm.

(2) There are strong reasons to believe that most (but not all) instances of human extinction would constitute a terrible tragedy.[1]

(3) Life-extension technologies could enable present or future people to live indefinitely long lives; some emerging technologies could also make life far more worth living than it currently is.

(5) These technologies could thus enable humanity to survive indefinitely with lives worth continuing.

(6) Therefore, one can coherently espouse (1) and reject Benatar's pro-extinction position if one also endorses the development of safe and effective life-extension technologies. Call this "no-extinction anti-natalism."

The following sections provide reasons for accepting these premises. Section 2 recapitulates the anti-natalist arguments put forward by Benatar. Section 3 adumbrates multiple arguments that independently converge upon the conclusion that one ought to oppose human extinction. Section 4 examines the question of whether humanity would be ethically justified in continuing to procreate until life-extension technologies become available. Section 5 explores the implications of no-extinction anti-natalism with respect to the metaphysics of diachronic personal identity; this will focus on mind-uploading, the duplication of uploaded minds, and the possibility of radical cognitive-psychological enhancement, in particular. Finally, section 6 concludes the paper.

Two thoughts before moving on to the substance of this essay. First, the argument here presented could very well *increase* the general appeal of anti-natalism among both academics and the public. As Benatar observes, many people intuitively accept the "starting point" of his argument, i.e., the harm-benefit asymmetry, yet spurn its apparent entailment that humanity

should go extinct, and hence reject anti-natalism. For people who, like myself, are sympathetic with the moral prescription not to procreate but are also inclined to think that human extinction would constitute an immense tragedy, the present paper offers a middle path that enables one to have one's cake and eat it, too. A lifeless universe would have been best, but a universe that contains intelligent beings that go extinct prematurely is worse than one that contains intelligent beings indefinitely. Second, on a personal note, I am not entirely convinced of the conclusions here outlined. Nonetheless, I endorse an epistemic distinction between *ideas worth considering* and *ideas worth accepting*. The position articulated by (5) above constitutes at least the former, and hence this paper aims at minimum to fill-in a lacuna in the literature on anti-natalism, thereby contributing to future discussions of the topic.

2. Benatarian Pro-Anti-Natalist Arguments

Philosophers have advanced a number of arguments for anti-natalism that proceed from different philosophical starting points. For example, Gerald Harrison (2012) proposes a deontological argument for anti-natalism based on W.D. Ross's notion of a "*prima facie* duty," that is, "a type of action that has a tendency to be right, other things being equal" (Ross 1998). Duties of this sort are analogous to "forces" that push and pull in sometimes opposing directions; when an opposing force—a conflicting *prima facie* duty—is absent, "then the action is pulled all the way to rightness or wrongness" (Harrison 2012, 96). Harrison argues that humans have a *prima facie* duty not to create people who will suffer, since we have a *prima facie* duty to prevent suffering. But we have no *prima facie* duty to create more happy people, since having a duty to perform an action A requires that not doing A would wrong someone. And since not creating a person P does

not wrong P, there is no *prima facie* duty to create P. This implies that we have no positive duty to procreate; but does it also imply that procreation is wrong? The answer is "yes" because if life contains both pleasures and pains, and if one has a *prima facie* duty to avoid pains but not to realize pleasures, then there are no reasons to create P and one reason not to create P, even if P would have a very happy life. Consequently, as Harrison puts it, "the *prima facie* duty to prevent the suffering is unopposed and thus decisive" (Harrison 2012, 98).

Along different lines, Asheel Singh (2012) draws from Seana Shiffrin's 1999 paper "Wrongful Life, Procreative Responsibility, and the Significance of Harm" in outlining a rights-based defense of anti-natalism. This begins with the claim that harming someone without their consent is wrong because it constitutes an infringement of her or his rights. Since it is impossible to ask an unborn person whether she or he would consent to being born, procreation is morally wrong. One response is that causing harm is not wrong if one has a reasonable expectation that the harm caused will preclude an even greater harm, as in the hypothetical case of Philippa pushing Derek into a ditch to prevent Derek from being hit by a bus. Derek might then say to Philippa once he gets back on his feet, "I'm so *glad* that you did that!" Applying this to procreation, many people aver to be glad that they exist—they *endorse* the pro-natalist decisions of their parents. Singh thus calls this the "endorsement objection" to his (and Shiffrin's) argument. But he offers a variety of reasons for rejecting this objection, which I will not recapitulate. Suffice it to say that, if Singh is successful, it could be that procreation violates the rights of people created and is therefore immoral.

However, this paper will focus almost exclusively on the consequentialist arguments of David Benatar in his 2006 book *Better Never to Have Been* (see also Benatar 1997). These take two general forms, which Benatar classifies as *philanthropic* and *misanthropic*, the former of

which subsumes both a philosophical and empirical argument.[2] Considering the philanthropic arguments in turn: the first case hinges on an asymmetry between pain (harms) and pleasure (benefits). The presence of pain is uncontroversially bad and the presence of pleasure is uncontroversially good. Less obvious is the additional assertion that "the absence of pain is good, even if that good is not enjoyed by anyone, whereas the absence of pleasure is not bad unless there is somebody for whom this absence is a deprivation" (Benatar 2006, 30)[3] Hence, a lack of pleasure is bad only if an existing person is deprived of it; otherwise this is "not good, but not bad either." In contrast, a lack of pain is good whether or not there exists a person to experience the state of lacking. This is the *harm-benefit asymmetry*, and Benatar argues that it can explain four other asymmetries that are "widely endorsed." He thus claims that the harm-benefit asymmetry may already be "widely accepted," even if most people do not realize this. These additional asymmetries are:

> (i) The intuition known as the "procreation asymmetry." This arises from the belief that we have a duty not to bring into existence people who will suffer, but no corresponding duty to create new people who will be happy. According to Benatar, we have this intuition because a life worth living is good for those who exist but not bad for those who don't exist—since, once again, in the latter case there is no one who is deprived of this worthwhile life. But the presence of a miserable life is bad, while the absence of a miserable life is good, even though no one exists in the latter case to experience this absence.
>
> (ii) The intuition that "whereas it is strange (if not incoherent) to give as a reason for having a child that the child one has will thereby be benefited," but "it is not strange to cite a potential child's interests as a basis for avoiding bringing a child into existence."

(iii) The intuition that "only bringing people into existence can be regretted *for* the sake of the person whose existence was contingent on our decision."

(iv) The intuition that discovering that a habitable island or exoplanet is uninhabited by happy people does not elicit the same degree of sadness as discovering that an island or exoplanet is populated by people who suffer greatly (Benatar 2006, 31-35).

Insofar as the harm-benefit asymmetry can account for these common views, Benatar argues that it should "have broad appeal" (Benatar 2006, 36). Yet if one accepts the harm-benefit asymmetry, then one must also accept the anti-natalist contention that bringing someone into existence is always a net harm. Consider the following two scenarios called "Existing" (X) and "Not-Existing" (~X). In X, the existence of an individual entails the presence of both pain and pleasure, the former being bad and the latter being good. In ~X, the non-existence of an individual entails the absence of both pain and pleasure, the former being good and the latter being not bad (or good). Now ask which of these scenarios is morally better. Since X yields a situation that is *good and bad* while ~X yields one that is *good and not bad*, it appears that ~X is better than X, from which it follows that non-existence is always preferable to existence.

The second, more empirical argument aims to show that our lives are full of (a) great suffering, and (b) more suffering than most of us realize. The latter is due, Benatar argues, to psychological distortions like the optimism bias and habituation, as well as our tendency to judge life by relative rather than absolute criteria: "A's life is *better than* B's, therefore A's life must be *good*." This makes self-reports about one's life-quality unreliable. As Benatar writes, "if people realized just how bad their lives were, they might grant that their coming into existence was a

harm even if they deny that coming into existence would have been a harm had their lives contained but the smallest amount of bad" (Benatar 2006, 60).

Consider that there have existed roughly 113 billion humans so far, meaning that some 105.4 billion people have already made the journey from the cradle to the grave (given the current population of 7.6 billion) (Kaneda and Haub 2018). This Kilimanjaro of death is the result of murders, suicides, genocides, wars, accidents, disease, and aging. More specifically, 15 million people may have perished in natural disasters in the last millennium; some 840 million people suffer from malnutrition and hunger, and roughly 20,000 die every single day from the latter; 133 million people may have died in mass killings before the twentieth century; some 110 million humans perished in the hemoclysmic conflicts of the twentieth century; and infectious/chronic diseases continue to trip millions into the eternal grave each year (see Benatar 2006). Elsewhere, I have calculated that the Black Death may have eliminated more people than World War II, the Taiping Rebellion, Mongol conquests, World War I, Napoleonic Wars, Vietnam War, American Civil War, 2003 Iraq War, and the War of 1812 *combined*—and public health experts claim that "we are at greater risk than ever of experiencing large-scale outbreaks and global pandemics," where "the next outbreak contender will most likely be a surprise" (see author; Senthilingam 2017).[4] Hence, the amount of human suffering throughout history is utterly inscrutable.[5] These are the two philanthropic arguments; I will address the misanthropic argument in section 6.

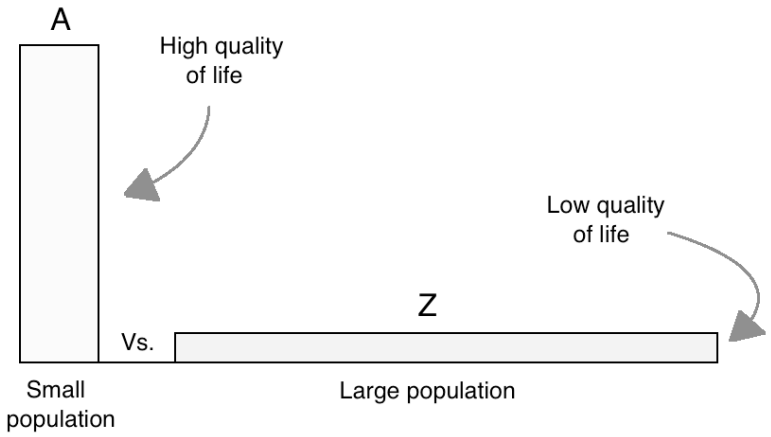
But there are additional reasons to accept Benatar's anti-natalism: his theory also appears to solve a number of recalcitrant problems in population ethics. For example, it seems to imply that a "narrow" version of the person-affecting view could, in fact, solve the "non-identity problem." This problem arises in situations where an action appears to wrong someone without harming that person, because the action changes that person's identity. Thus, if by "harm" one means

"making someone worse off than she otherwise would have been," then a wide range of intuitively wrong actions appear permissible.[6] As for the narrow person-affecting view, Parfit delineates it as follows:

> Suppose that we are comparing outcomes *X* and *Y*. Call the people who will exist in outcome X *the X-people*. Of these two outcomes, call X 'worse for people' in the narrow sense if the occurrence of X rather than Y would be either worse for, or bad for, the X-people (Parfit 1984, 398).

If we stipulate that X is "coming into existence" and Y is "not coming into existence," then X is "worse for people" than Y, given the harm-benefit asymmetry. Does this entail comparing two states of the same person, one in which that person does not exist? No, according to Benatar: "What we mean when we say that somebody would have been better off not having come into existence is that non-existence would have been preferable." In other words, the comparison is between the existence of P and "an alternative state of affairs in which [P] does not exist" (Benatar 2006, 22).[7] This solves the non-identity problem because the identity of the person brought into existence is irrelevant: in all cases, creating that person is wrong.

Benatar claims that his view also sidesteps the *Repugnant Conclusion* and *mere addition paradox* that undercut total and average utilitarianism. Since smaller populations are always preferable to larger populations, this (a) makes Parfit's "population Z" (consisting of many people with very low-quality lives) worse than "population A" (consisting of few people with very high-quality lives), and (b) renders *any* addition of people—whether their quality of life is higher *or* lower than the average—morally unacceptable. (See figure 1.)



Thus, there are cogent arguments for propositions (i) and (ii) in section 1: it would have been better if there were no people at all, and it would be good if humanity desists from creating new people. Benatarian anti-natalism also appears to circumvent the most intractable conundrums of population axiology—all of which suggest that Benatar is on to something.

3. Until Entropy Death Do Us Part?

Yet few academics, not to mention the general public, would self-identify as "Benatarians." The most obvious reason pertains to the default biological impulse to reproduce. By analogy, I know that it is wrong (for my health) to eat junk food, but the selective environment in which my species evolved was one in which salt was not abundant; hence, I have a strong urge to consume salt. Another unsavory feature for many people is the implication that humanity should disappear, sooner rather than later, by voluntarily giving up procreation. As Benatar writes, his philanthropic arguments "imply that it would be better if humans (and other species) became extinct," adding that "the case for earlier extinction is … strong" (Benatar 2006, 194, 198). Benatar refers to the process of ending human life by this means as a "dying-extinction," contrasting it

with scenarios in which humanity is *killed off* by natural or anthropogenic factors, which he terms a "killing-extinction." Crucially, Benatar seems to believe that his pro-extinction view directly and obviously follows from his central thesis "that there should, ideally, be no (more) people." In his words,

> my answer to the question "How many people should there be?" is "zero." That is to say, I do not think that there should ever have been any people. Given that there have been people, I do not think that there should be any more (Benatar 2006, 182).

But claiming that there should not be any more people does *not* imply that the population should be zero. As already stated: life-extension technologies could enable people to live indefinitely long lives—that is, to attain what I call "functional immortality," whereby one lives until she either succumbs to injury, commits suicide, or perishes due to the heat death of the universe.[8] This gestures the "dual policy" at the heart of no-extinction anti-natalism: first, we should stop procreating entirely (anti-natalism), and second, we should develop safe and effective life-extension technologies (pro-immortalism). This should appeal to Benatar himself given his anti-Epicurean view that death is (often) bad and, as such, one ought to (typically) avoid it. The reason is that, when evaluating an outcome, *any* harm at all "will be decisive." Since "we (usually) have an interest in continuing to exist … death may be thought of as a harm, even though coming into existence is also a harm" (Benatar 2006, 213). This links to an important distinction between "lives worth starting" and "lives worth continuing." Although life is much worse than most of us realize, many lives aren't *so bad* as to warrant suicide. Thus, they are worth continuing even though they weren't worth starting, much like a movie that is worth finishing but not having started. (If

only you had known the movie was bad, you would have stayed home rather than venturing out to the theater.)

But just as there are human *lives* worth continuing, there are arguments for why human *existence* is worth continuing, even if it would have been better for our evolutionary lineage never to have been. The strongest argument against human extinction, if sound, comes from total utilitarianism. The first to articulate this idea was Henry Sidgwick (1874), although more recent versions proceed by calculating how many people could come to populate Earth or our future light cone, with some claiming that up to 10^{46} "human lifetimes" could exist per century within our galactic supercluster (Ćirković 2002). Since people are the substrate or "containers" of well-being (or value), the more people there are with net positive amounts of well-being, the better things will go from the "point of view of the universe." But creating new lives is incompatible with anti-natalism, so this line of argumentation is a non-starter in the present context. There is, however, an alternative approach that focuses not on the total number of people but on the *quality* of their lives—rather than merely "lives not so bad as to warrant suicide," there could exist future life conditions that are *extremely good* and hence *very much* worth continuing.

Why believe that life in the future could improve? There are at least two reasons. First, there is at least some evidence that violence has declined over the past several centuries, if not longer (see Ferguson 2012 for criticisms). Since the middle of the twentieth century, the rights of many groups has expanded in the West, including women, minorities, LGBTQ people, and non-human animals—a process driven by what Steven Pinker names the "moral Flynn effect" (Pinker 2011). If this trend continues, we should expect the world to become even more peaceable and tolerant in the future; as Martin Luther King Jr. hypothesized, "the arc of the moral universe is long, but it bends toward justice." Second, some transhumanists argue that radical human en-

hancements could significantly improve the human condition by augmenting core human capacities like cognition, emotion, wisdom, and morality (Bostrom 2008; Persson and Savulescu 2012). Transhumanist philosophers tell us that such enhancements could create a *posthuman* condition that is far superior in uncontroversial ways than the present *human* condition, which is marked by pervasive ignorance, violence, foolishness, and iniquity. Together, these two trends could greatly decrease the "magnitude," to borrow Benatar's term, of harm in the world. In doing so, they would yield a future situation marked by far more well-being than there is today, which is good from a total (and average) utilitarian perspective.

But one need not embrace utilitarianism to strongly oppose human extinction. I have elsewhere explored these arguments in great detail, so I will here offer only some brief summaries, directing readers to my prior work (see author and co-author, forthcoming). The point is to underline that there are multiple lines of argumentation that (a) are compatible with Benatarian anti-natalism, and (b) independently converge on the conclusion that human extinction, if it were to occur, would constitute an immense tragedy. A few of these are:

(i) *The argument from unfinished business*. This states that, to quote Edmund Burke (1790), civilization is "a partnership not only between those who are living, but between those who are living, those who are dead, and those who are to be born." That is to say, humanity is "building something" over time—a morally better society, a complete scientific picture of the universe, a world with no more scarcity, and so on—such that it behooves us not to end up being the failing link in the chain of generations.

(ii) *The argument from cosmic uniqueness*. According to a recent study that replaces the variables in the Drake equation with probability distributions, there is a very high proba-

bility that we are the only intelligent lifeforms in the Milky Way galaxy, if not the visible universe (Sandberg et al. 2018). To quote Parfit (2011) once more:

> if we are the only rational beings in the Universe … it matters even more whether we shall have descendants or successors during the billions of years in which that would be possible. Some of our successors might live lives and create worlds that … would give us all … reasons to be glad that the Universe exists.[9]

(iii) *The argument from normative uncertainty*. There could be reasons, whether distinctly moral or not, for maintaining that human extinction would be extremely undesirable, but that the philosophical enterprise from the pre-Socratics to the present has not yet discovered. This gives us reason to hope that humanity does not go extinct, even if we believe that it should or that extinction would not ultimately be that bad. Tomorrow, a young genius could flip over a stone to find a novel insight that "radically changes the expected value of pursuing some high-level subgoal" (Bostrom 2014).[10]

(iv) *The argument from value-ladenness*. Samuel Scheffler (2016, 2018) argues that most of us care about the continuation of our evolutionary lineage far more than we realize—that we have a "love for humanity." This is supposedly revealed by the putative fact that most of us would become despondent if told, with certitude, that humanity would go extinct several decades after our deaths. As Scheffler writes, referring to future generations, "we have an interest in their survival in part because they matter to us; they do not matter to us solely because we have an interest in their survival" (Scheffler 2018).

(v) *The argument from final value*. As Johann Frick (2017) observes, people commonly attribute "final value" (when something is valuable for its own sake) to a range of phenomena like languages, species, and cultures. This suggests, he argues, that humanity,

> "with its unique capacities for complex language use and rational thought, its sensitivity to moral reasons, its ability to produce and appreciate art, music, and scientific knowledge, its sense of history, and so on, should be deemed to possess final value" (Frick 2017, 359). Since it would be nonsensical to value things but "see no reason of any kind to sustain them or retain them or preserve them or extend them into the future" (Scheffler 2007, 106), we should conclude that "it would be very bad, indeed one of the worst things that could possibly happen, if, for preventable reasons, the end came much sooner rather than later" (Frick 2017, 344).
>
> (vi) *The argument from harm reduction in the wild*. Some argue that there exists far more pain than pleasure in the natural world, in part because of *r*-selected species that give birth to tens or hundreds of offspring each reproductive cycle, only a relative few of which survive. If humanity exists long enough into the future, it could potentially develop ways to intervene in the ecosystem hierarchy to reduce the magnitude of harm in the biosphere.

Many more arguments for ensuring our survival could be adduced. In contrast, there are few compelling reasons for either indifference toward human extinction or seeing this outcome as desirable. Perhaps the two most notable are the *argument from negative utilitarianism* and the *argument from anthropogenic destruction* (which ties into Benatar's misanthropic case for anti-natalism). The first claims that all that matters morally is the total amount of suffering in the world, independent of the total amount of well-being. Thus, if possible, one should become a "world-exploder" who destroys not just all sentient life on Earth, but the potential for sentient life to exist in our future light cone. (One realistic way of doing this might be to weaponize a par-

ticle accelerator. *If* the universe is in a "false vacuum" state, then it is theoretically possible to tip it into a "true vacuum" state by nucleating a "vacuum bubble" that expands in all directions at the speed of light, obliterating everything it encounters.[11]) The second claims that humans have destroyed so much of the biosphere that we deserve to perish. There are several fringe groups motivated by this view, such as the Gaia Liberation Front, that advocate for the involuntary annihilation of humanity; other groups, such as the Voluntary Human Extinction Movement (VHEMT), argue that humanity should voluntarily cease procreating for the sake of the "Gaian system." (Thus, VHEMT are anti-natalists but for reasons unrelated to the harm-benefit asymmetry.)

Considering these for and against arguments together, there is, I believe, overwhelming reason to advocate for the continued existence of humanity. To quote Derek Parfit (1984),

> civilization began only a few thousand years ago. If we do not destroy mankind, these few thousand years may be only a tiny fraction of the whole of civilized human history. The difference between [nearly all and actually all people dying] may thus be the difference between this tiny fraction and all of the rest of this history. If we compare this possible history to a day, what has occurred so far is only a fraction of a second.

Since there are reasons for not wanting the human population to fall to zero, but also reasons for not creating more people, one is left with the no-extinction anti-natalist position that we should stop creating more people while actively promoting the development of safe and effective life-extension technologies. But this view encounters a number of serious problems of its own. The following two sections will examine a handful of these problems: first, problems relating to the

fact that life-extension technologies are not yet available, and second, complications associated with functional immortality and diachronic personal identity.

Section 4: Could No-Extinction Anti-Natalism Work?

There are at least two ways that functional immortality could be achieved. One is called "whole-brain emulation" or "mind-uploading," which involves simulating the microstructure of the brain on a computer with sufficient fidelity to reproduce consciousness. This assumes that minds are multiply realizable, by virtue of being "organizational invariants" that arise from systems exhibiting the right functional organization, whatever the physical substrate (Chalmers 1996). If this "can be made to work," it would constitute "the ultimate life-extension technology," since uploaded minds "would not be subject to biological senescence" and "back-up copies could be created regularly so that you could be re-booted if something bad happened (thus your lifespan would potentially be as long as the universe's)" (Labrecque 2017, 166). There are three main forms of uploading: (i) *destructive uploading* (the original brain is destroyed either gradually or instantaneously), (ii) *non-destructive uploading* (the original brain remains fully intact while a copy is made), and (iii) *reconstructive uploading* (the original brain dies but the person is recreated based on historical records) (Chalmers 2010; see also Sandberg and Bostrom 2008).

One version of destructive uploading is the "microtome procedure." This involves freezing a recently deceased brain to liquid nitrogen temperatures, slicing it into small sections, scanning the slices, transferring this information to a computer, and then simulating the brain. A non-destructive option is to scan the brain "from within … using nanobots" that transfer this informa-

tion via wireless systems to a computer (Kurzweil 2005). A reconstructive option could become possible if, for instance, a future superintelligent machine were to collect enough data about a past person to design a program that instantiates the relevant functional-organizational properties of that person's brain (Chalmers 2010). This is predicated on a distinction between *clinical death* and *information-theoretic death*, where the latter refers to the point at which no information about one's nervous system is permanently lost (Merkle 2018).

The alternative way to extend life is through biomedical anti-aging interventions. For example, *geroprotectors* are drugs that decelerate aging; the anti-diabetes drug metformin, for example, has been shown to reduce "all-cause mortality and diseases of ageing independent of its effect on diabetes control" (Campbell 2017). And lifelong *caloric restriction* has been observed to "considerably [extend] both the healthy and total life span of nearly all species in which it has been tried, including rodents and dogs" (de Grey 2004, 724). A more speculative possibility involves what Aubrey de Grey calls "strategies for engineered negligible senescence" (SENS). There are nine primary types of deleterious, cumulative changes that are associated with aging: cell loss (without replacement); oncogenic nuclear mutations and epimutations; cell senescence; mitochondrial mutations; lysosomal aggregates; extracellular aggregates; random extracellular protein cross-linking; immune system decline; and endocrine changes (de Grey 2003). Interventions to halt or reverse these changes using CRISPR/Cas9 and other synthetic biology systems could thus potentially halt or reverse senescence itself.

The question is whether any of these technologies will become available before humanity dies out, if anti-natalist policies were implemented today. According to Ray Kurzweil (2005), the nanobotic technology required to scan the brain "will be available by the late 2020s" and the "computational performance, memory, and brain-scanning prerequisites of uploading" will

emerge in "the early 2030s." He surmises that "like any other technology, it will take some iterative refinement to perfect this capability, so the end of the 2030s is a conservative projection for successful uploading" (Kurzweil 2005). Geroprotectors and caloric restriction diets are already around, but might only add a few extra years to one's life. For example, de Grey notes that it could be possible for pharmaceuticals to "elicit the gene expression changes that result from caloric restriction," which might "extend human life span by something approaching the same proportion as seen in rodents—20% is often predicted—without impacting quality of life, and even when administered starting in middle age" (de Grey 2004, 724).

But it is important to note that a full-blown senescence-stopping technology need not be realized for one to attain functional immortality, meaning that even a few extra years could be extremely valuable. The reason concerns the idea of "actuarial escape velocity" (AEV), which refers to the fuzzily defined moment when advances in new anti-aging interventions occur faster than one ages using previous anti-aging interventions. Hence, one must only "live long enough to live forever" (Kurzweil 2004). De Grey elaborates this as follows:

> Those who get first-generation therapies only just in time will in fact be unlikely to live more than 20-30 years more than their parents, because they will spend many frail years with a short remaining life expectancy (i.e., a high risk of imminent death), whereas those only a little younger will never get that frail and will spend rather few years even in biological middle age. Quantitatively, what this means is that if a 10% per year decline of mortality rates at all ages is achieved and sustained indefinitely, then the first 1,000-year-old is probably only 5-10 years younger than the first 150-year-old (de Grey 2004, 725).

So, the revised question is when humanity will cross the threshold of AEV. On de Grey's view, the first people to become functionally immortal may have already been born. In his words, "if we ask the question: 'Has the person been born who will be able to escape the ill health of old age indefinitely?' Then I would say the chances of that are very high … Probably about 80 percent" (quoted in Jolliffe 2015). Similarly, Kurzweil prognosticates that "we will reach a point around 2029 when medical technologies will add one additional year every year to your life expectancy … By that I don't mean life expectancy based on your birthdate, but rather your remaining life expectancy" (quoted in Ranj 2016). If this sounds too optimistic, it may be worth recalling Nicholas Wade's (2004) observation that major advances in biotechnology—such as sequencing the human genome—have often occurred *sooner* than experts anticipated. Thus, Wade speculates that "longevity increases might be one of those big steps that arrive much sooner than expected."

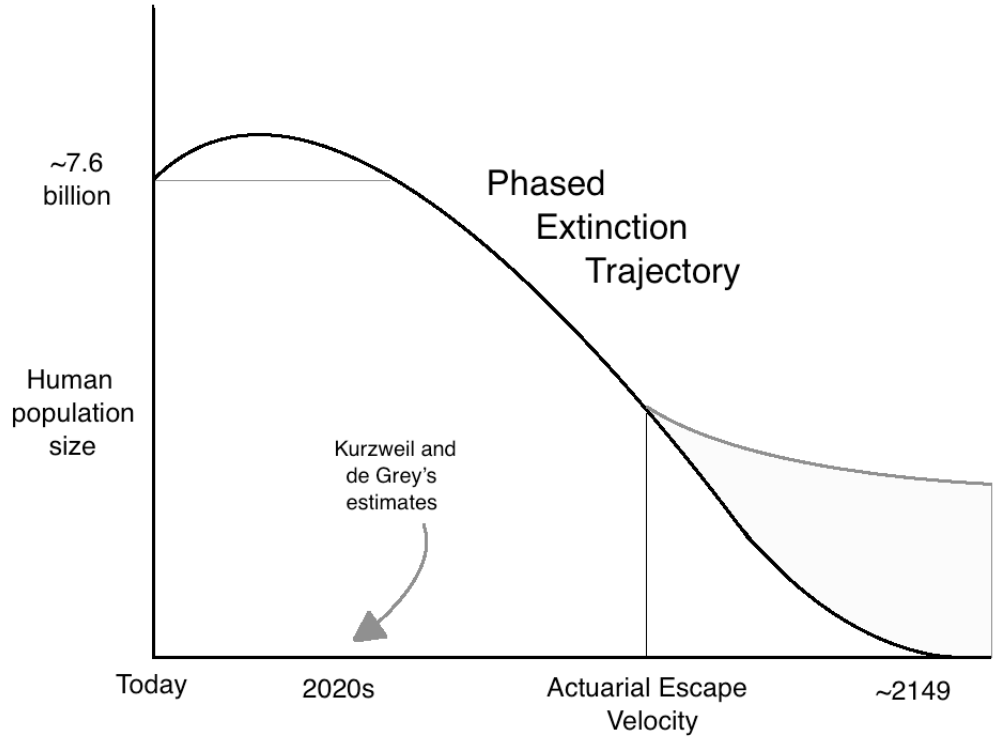
These estimates suggest that the global population could stop procreating today without humanity going extinct. But how reliable are Kurzweil's and de Grey's prognostications? Who knows: after all, studies show that at least some classes of experts are no better at predicting the future than non-experts, and to my knowledge neither Kurzweil nor de Grey are "superforecasters" (Armstrong and Sotala 2015; see Tetlock and Gardner 2015). Thus, let's say that these estimates are far from the mark. Let's say that AEV won't arrive until 2110, when most people born today will have died, given an average lifespan of ~80 years. Does this mean humanity is doomed if all people stop procreating tomorrow? Or does anti-natalism have the resources to justify creating new people until we cross the finish line of AEV?

Consider Benatar's argument that lives in the final generation before extinction could be so impoverished that there may be reason to "phase-out" humanity rather than to stop procreating

all at once; call this *humane extinction*. Benatar identifies two moral positions that support this idea: (1) a "negative total view" according to which "we may create new people where the total amount of harm in doing so is equivalent to, or less than, the harm that would be suffered by existing people if the new people were not created." And (2) a "less stringent rights or deontological view" according to which "creating new people cannot be justified by mere reduction in total harm," but it "may be justified by substantial (but not mere) reduction in total harm" (Benatar 2006, 192). (The narrow person-affecting view to which Benatar previously appeals to dodge the non-identity problem cannot confer moral license to procreate under any circumstances.) Benatar thus concludes that his "views might allow, under some conditions, for a phased extinction, whereby fewer and fewer children are brought into existence in each of (only) a few successive generations, rather than an immediate cessation of all baby-making" (Benatar 2006, 17).

The question is thus whether the negative total or the rights or deontological view could justify further procreation to avoid human extinction. Unfortunately, the answer is "no," since both views focus entirely, in Benatar's formulation, on reducing harm to and only to *extant* people. If creating new people doesn't reduce harm to those already alive, then procreating is unjustified. But one could exploit this line of reasoning as a work-around to the problem of creating new people to reach AEV, if that were to be necessary. Consider the following statement from Benatar:

> Whether the number of people could be reduced fast enough without the costs of rapid population decline, to a level where the number of final people was small enough to offset the harm to intervening generations, is a difficult one to answer. Whatever the answer,



> we can say that extinction *within a few generations* is to be preferred to extinction only after innumerably more generations (Benatar 2006, 198; italics added).

Benatar thus suggests that continuing to procreate for a "few" more generations could be permissible. If we interpret "few" as denoting "3," if the length of a contemporary generation is ~25.5 years, and if the average lifespan remains at ~80 years, then the last humans on a phased-extinction schedule would die in about 131 years, circa 2150. In other words, children born today would have children in 25.5 years, these children of children would have children 25.5 years later, and these children of children of children would refrain from having children but live another 80 years.[12] Here I am conservatively bracketing the possibility of geroprotectors and dietary modifications extending the average lifespan by years or even decades.

The point is that *even if* the estimates from Kurzweil and de Grey are way off the mark, there is reason for optimism that humanity will cross the AEV threshold *even if* this requires surviving halfway into the twenty-second century. Although continuing to procreate *simply to reach* AEV may not be licensed by either the total negative view or the rights/deontological view, implementing a phased extinction policy *a la* Benatar's proposal could provide plenty enough wiggle-room to avoid extinction, i.e., without violating any Benatarian prescriptions, humanity could live long enough to live forever. From an impersonal perspective, this would be good insofar as one finds the anti-extinction arguments above good; from a prudential perspective, this would be good insofar as one accepts Benatar's anti-Epicurean position that death is (usually) bad.

Incidentally, there may be an argument based in Benatar's position for why procreating *until we reach AEV* might be justified. Take the following scenario: a child, C1, is diagnosed with leukemia and thus requires a bone marrow transplant to survive. The parents of C1 decide to have another child, C2, to enable C1's treatment. (The fetus of C2 could be screened to ensure that it's human leukocyte antigens match C1's.) Consequently, the parents treat C2 as a mere means to an end rather than an end in-itself, which of course violates the Kantian imperative, specified by his Formula of Humanity, to always treat people as ends and never as means. On Benatar's utilitarian view, when one considers the well-being of C2 alone, it is never morally permissible to bring C2 into existence if C2 will experience so much as a pin-prick. But the benefit of bringing C2 into existence for C1 would at least render procreation *more justified*. As Benatar writes, "it certainly seems strange to think that it is acceptable to have a child for no reason at all, but wrong to have a child in order to save somebody's life" (Benatar 2006, 130-131) Similarly, Benatar's phased extinction proposal rests on treating newly created people as a means to

an end: they serve the instrumental function of softening the hardship of being among the last people on Earth.

This suggests that if the value of an end is *sufficiently great*, then it could justify the means for achieving it. Hence, if an anti-natalist finds the anti-extinction arguments delineated above *extremely compelling*, then she may judge the benefit of creating new people as outweighing the harm caused by procreation. For space reasons, I will not discuss this argument any further, although I believe it is worth pursuing.[13]

Section 5: Personal Identity, Mind-Uploading, and Radical Human Enhancement

We have now offered reasons for accepting anti-natalism and opposing human extinction, and shown that no-extinction anti-natalism could work. But this view encounters a number of serious problems on its own, associated with (i) the possibility of attaining functional immortality through mind-uploading, and (ii) the likelihood that people who attain functional immortality will almost certainly modify their cognitive-psychological properties in radical ways, given enough time. The present section will focus on (i) and (ii), mostly ignoring the various more general problems with extending human lifetimes indefinitely.[14]

(i) For mind-uploading to provide a form of *digital* immortality, the uploaded person will need to be identical to the person uploaded. We can follow John Locke (1689) in defining a person (or self) as "a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places." The question is whether *people* are organizational invariants like minds, if minds are. That is, *does mind-uploading entail person-uploading*, or are these two distinct phenomena that can be de-coupled, if only asymmet-

rically? (For person-uploading to work, surely mind-uploading must work, too, but person-uploading need not work for mind-uploading to work.) The most clearly problematic case involves non-destructive uploading, since this would yield two numerically distinct persons who are psychologically continuous with a single original (who continues to exist as one of these beings). This leads to two issues:

First, if one maintains that persons, in the metaphysical sense, are intrinsically singular, then they cannot be organizational invariants, which means that the uploaded mind would in fact be a different *person* than the original; hence, non-destructive uploading would not enable one to achieve functional immortality. Why think that selves are intrinsically singular? Consider a case in which you visit a Digital Immorality Clinic. They pump your blood full of wirelessly connected nanobots that cross the blood-brain barrier and map-out the functional organization of your nervous system. This information is sent to a computer placed within a realistic android, similar to those in the TV show *West World*. You and this copy are later kidnapped and told that *either* you—the person that walked into the clinic—or your copy—the person that was created in the clinic—will be tortured for 48 hours and then murdered. Let's say care most about your own well-being. No rational person would then say, "Pick one of us, randomly. We're the same person, so it doesn't matter." To the contrary, someone acting out of prudence should exhort: "Torture the copy!" The idea that persons are singular entities makes sense of this situation.

Second, whether or not the uploaded mind is *you*, the fact is that post-upload there would exist two rather than just one conscious entity: the original and the functional isomorph. Hence, non-destructive uploading entails the creation of a new person, and since it is always wrong to create new persons, according to anti-natalism, it must always be wrong to non-destructively upload one's mind. Call this scenario "digital procreation." Anti-natalists should therefore *strongly*

oppose this form of uploading independent of one's view of personal identity. The same goes for reconstructive uploading: even if the functional isomorph is the same person, it would entail the creation of a new conscious being—a new being at T3 based on the informational patterns of a being who existed at T1 but died at T2—which would be wrong.

Destructive uploading is more complicated. Consider a piecemeal process whereby each biological neuron in one's brain is replaced by a functionally isomorphic artificial neuron. Mark Walker (2008) calls this "gradualism," and it seems intuitively plausible that it would preserve continuity of consciousness from the beginning to the end, at which point the whole brain has been replaced by non-biological matter. If continuity of consciousness is all that is needed to preserve personal identity, then the person at T2 will be the same as at T1. Benatarian anti-natalists should thus applaud this form of uploading, since it would enable the uploaded person to evade death. But if the person at T2 is different than at T1, the situation would be *doubly wrong*. The reason is that it would entail not only the creation of a new person, which would be bad, but the death of the original person, which would also (usually) be bad. But is continuity of consciousness sufficient for person-uploading? I agree with Chalmers (2010) that the metaphysics of diachronic personal identity are deeply confounding; they may very well be, in the sense of Colin McGinn (1993), permanently unknowable "mysteries" relative to the concept-generating mechanisms in our brains, the haphazard product of millions of years of contingent evolution.

Nonetheless, philosophers have defended various answers to this question. For example, Susan Schneider and Joe Corabi (2014) put forth a clever argument (not here recapitulated) for why destructive uploading would *not* preserve identity. As they write, if personal identity is simply a matter of continuity of consciousness, and if

> all that is required for this continuity is that later mental states be caused by or be qualitatively similar to earlier ones (e.g., that qualitatively similar thoughts be entertained, or that later sensory "memories" be qualitatively similar to the original experiences), then plainly the sort of continuity in question could be shared by numerous individuals at later times; what if the information were sent to two computers instead of just one, after all? (Schneider and Corabi 2014).

This relies on the aforementioned supposition that persons are singular entities. Massimo Pigliucci (2014) offers a different argument. Building on John Searle's "biological naturalism," he argues that *minds* are not organizational invariants, and that "functionalists" are guilty of embracing an "untenably strong" version of the computational theory of mind and misunderstanding the Church-Turing thesis. This problematizes person-uploading because while instantiating the same mind on a different substrate may not be sufficient to instantiate the same person, it does appear necessary.[15] (What would it mean for a person to become digital but not her or his mind?) Pigliucci thus concludes that mind-uploading is nothing more than "a very technologically sophisticated (and likely very, very expensive) form of suicide" (Pigliucci 2014, 129). Another argument comes from Nicholas Agar (2016), who claims that mind-uploading "does not satisfy a necessary condition for the transmission of human identities." The reason is that uploading constitutes a "significant interference in the properties and processes that normally accompany our survival." Consequently, "the replacement of human biology by machines may result in significant enhancement, but it will not result in significant enhancement for us" (Agar 2016, 184).

The point is not to claim that the evidence favors a pessimistic rather than optimistic view of mind-uploading with respect to personal identity. Rather, I merely wish to emphasize that

there is serious disagreement about the issue. And this should be enough for anti-natalists to be very cautious: if one accepts the harm-benefit asymmetry, and if death is (usually) bad, then even a low probability that destructive uploading kills the original while creating a new person should be sufficient for anti-natalists to *strongly oppose* gradualism. Furthermore, as Chalmers (2010) points out, if one is pessimistic (in terms of the preservation of personal identity) about gradualistic uploading, then one should be even more pessimistic about instantaneous uploading, whereby the wetware of the brain is abruptly replaced by non-biological hardware, rather than through a piecemeal process. Hence, one should be even more strongly opposed to instantaneousism (as we might call it).

A related problem concerns the possibility of *duplication* gestured at by Schneider and Corabi (2014). As they suggest, once a mind M is uploaded, it could easily be duplicated on multiple computers an indefinite number of times. This is orthogonal to the question of whether M would be personally identical to the original; what matters is that before M has been duplicated, the duplicate M' does not exist, while after M has been duplicated, both M and M' exist. Call this "duplicative procreation." It would thus entail the creation of a new conscious entity, thereby instantiating a state of affairs that is both good and bad, compared to the original situation that was good and not bad. This is not a trivial point: it directly connects to speculations outlined by futurists like Robin Hanson (2016). Referring to uploads as "ems," Hanson argues that, in a post-upload economy, ems could (or would, or should) create short-lived duplicates, called "spurs," to complete specific work-tasks, after which they would be promptly terminated. In his words,

> most ems … are comfortable with often splitting off a "spur" copy to do a several hour task and then end, or perhaps retire to a far slower speed. They see the choice to end a

> spur not as "Should I die?" but instead as "Do I want to remember this?" At any one time, most ems are spurs (Hanson 2016, 10).

This is a disturbing scenario for many ethical reasons. Since spurs are numerically separate entities, terminating them would be tantamount to murder—a type of "mind crime," in Bostrom's (2016) phraseology. Furthermore, as Anders Sandberg (2014) notes, "if ending the identifiable life of [a spur] is a wrong, then it might be possible to produce a large number of wrongs by repeatedly running and deleting instances of an emulation even if the experiences during the run are neutral or identical" (Sandberg 2014, 288).[16] Even more, since creating spurs entails duplicating mind-uploads, anti-natalists should be *especially* worried about this futuristic scenario obtaining, since it would mean both creating and killing a conscious entity.[17]

(ii) So far, the challenges discussed arise specifically from (a) modifications to the underlying substrate upon which the conscious mind supervenes, and (b) the possibility of copying minds once they are simulated by a computer. But similar problems emerge from the likelihood that functionally immortal persons will almost certainly alter their cognitive-psychological properties in radical ways given enough time. Consider the following hypothesis, which I believe is highly plausible:

> *Inevitable Modification Hypothesis (IMH)*: any population of intelligent beings with technology advanced enough to enable functional immortality will also eventually use related technologies to radically alter their cognitive-psychological properties.

This holds, I contend, whether we attain functional immortality through biomedical anti-aging interventions or mind-uploading. If we cross AEV as biological or technobiological hybrid beings, genetic engineering, nootropics, and brain-computer interfaces (BCIs) could radically enhance our core emotional and intellectual capacities. If we attain immortality by uploading our minds, a wide variety of speculative interventions could augment our capacities in significant ways. Perhaps uploaded minds could recursively self-improve, resulting in an "intelligence explosion" that yields intellects many orders of magnitude smarter than current humans.

The point is that if IMH is true, it (re)introduces similar personal identity problems to those discussed above. Let's call the unenhanced minds "M-" and the enhanced minds "M+." Will the transition from M- to M+ preserve identity? If so, Benatarians might advocate for radical cognitive-psychological enhancement, since M+ would presumably (if not by definition) have a life more worth continuing than M-. If not, then Benatarians should see this transition as *doubly wrong*, since M+ would entail the creation of a new person, in the metaphysical sense, and literal death of M-. Call this scenario "developmental procreation." Note that it differs from duplication as follows: duplication involves creating two numerically distinct but qualitatively identical entities (at the moment of duplication), whereas enhancement involves creating two qualitatively distinct but numerically identical entities.

The question now is whether we have good reason to think that M+ is the same as M-. Naturally, there is disagreement. For example, Mark Walker (2008) defends an "Aristotelian identity argument" according to which "some changes may be so drastic that they will mean that I cease to exist." To use Aristotle's own example, if a "man wishes his friend's good for his friend's sake, the friend would have to remain the man he was. Consequently, one will wish the greatest good for his friend as a human being" rather than him becoming a superhuman

"god" (Walker 2008). Schneider (2008) propounds an even stronger view according to which "even mild enhancements are death inducing." To quote her at length,

> for radical enhancement to be a worthwhile option for you, it has to represent a desirable form of personal development; at bare minimum, even if enhancement brings such goodies as superhuman intelligence and radical life extension, it must not involve the elimination of one of your essential properties. *For in this case, the sharper mind and fitter body would not be experienced by you—they would be experienced by someone else*. For even if you would like to become superintelligent, knowingly embarking upon a path that trades away one or more of your essential properties would be tantamount to suicide—that is, to your intentionally causing yourself to cease to exist (Schneider 2008, 5).

Once again, the fact that no consensus on this topic exists should lead no-extinction anti-natalists to *strongly oppose* radical modifications to one's cognitive-psychological properties.[18]

6. Conclusion

This paper has outlined a novel interpretation of anti-natalism according to which Benatarians can have their cake and eat it, too—as the cliche goes. I have claimed that one can coherently accept the harm-benefit asymmetry, maintain that there should be no more people, and advocate the continued indefinite survival of humanity. Indeed, there is a range of independently strong arguments that all converge upon the conclusion that human extinction—the process of

going extinct, if not the condition of being extinct—would constitute an immense tragedy. I then examined a number of potential problems for no-extinction anti-natalism.

Finally, we should note here that no-extinction anti-natalism does not address the various misanthropic arguments that Benatar propounds. Here one could retort that, if those living in the future become radically enhanced posthumans, then the reasons for disliking humans may very well no longer apply. As alluded to above, it could be possible for a combination of cognitive enhancements and moral bioenhancements—such as *mostropics*, coined on the model of *nootropics*—to significantly boost our individual and collective wisdom, morality, compassion, sympathy, kindness, and so on (see Persson and Savulescu 2012; author 2016). If this quasi-utopian possibility were to obtain, then the "superb misanthropic argument against having children and in favour of human extinction" that "rests on the indisputable premiss that humans cause colossal amounts of suffering" could become irrelevant (Benatar 2006, 244). The view here outlined thus offers at least some hope of overcoming the misanthropic argument for favoring human extinction, although, as explored in the previous section, problems and puzzles abound.

Although Benatar does not endorse *pro-mortalism*, he does oppose *pro-immortalism*. But this need not be the case. While I am not fully convinced that no-extinction anti-natalism is the best moral view, I believe this position deserves a hearing in the court of philosophical disputation.

References:

Adams, Fred. 2008. Long-Term Astrophysical Processes. In Nick Bostrom and Milan Ćirković (eds.), *Global Catastrophic Risks*. Oxford: Oxford University Press.

Adams, Robert. 1989. Should ethics be more impersonal? A critical notice of Derek Parfit, Reasons and persons, *Philosophical Review*. 98 4): 439-484.

Agar, Nicholas. 2016. Enhancement, Mind-Uploading, and Personal Identity. In Steve Clarke, Julian Savulescu, C.A.J. Coady, Alberto Giubilini, and Sagar Sanyal (eds.) *The Ethics of Human Enhancement: Understanding the Debate*. Oxford: Oxford University Press.

Althaus, David, and Lukas Gloor. 2018. Reducing Risks of Astronomical Suffering: A Neglected Priority. Foundational Research Institute. https://foundational-research.org/reducing-risks-of-astronomical-suffering-a-neglected-priority/.

Bell, Wendell. 1993. Why Should We Care About Future Generations? In H.F. Didsbury, Jr. (ed.) *The Years Ahead: Perils, Progress, and Promises*. Bethesda, MD: The World Future Society.

Benatar, David. 1997. Why it is Better Never to Come into Existence. *American Philosophical Quarterly*. 34: 345-355.

Benatar, David. 2006. *Better Never To Have Been: The Harm of Coming Into Existence*. Oxford: Oxford University Press.

Bostrom, Nick. 2003a. The Transhumanist FAQ: A General Introduction, Version 2.1. https://nickbostrom.com/views/transhumanist.pdf.

Bostrom, Nick. 2003b. Are You Living in a Computer Simulation? *Philosophical Quarterly*. 53(211): 243-255.

Bostrom, Nick. 2008a. Three Ways to Advance Science. *Nature* Podcast. https://nickbostrom.-com/views/science.pdf.

Bostrom, Nick. 2008b. Why I Want to be a Posthuman When I Grow Up. In Bert Gordijn and Ruth Chadwick (eds.) *Medical Enhancement and Posthumanity*. New York, NY: Springer.

Bostrom, Nick. 2013. Existential Risk Prevention as Global Priority. *Global Policy*. 4(1): 15-31.

Bostrom, Nick. 2014a. Crucial Considerations and Wise Philanthropy. Soundcloud. https://soundcloud.com/gooddoneright/nick-bostrom-crucial-considerations-and-wise-philanthropy.

Bostrom, Nick. 2014b. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bratsberg, Bernt, and Ole Rogeberg. 2018. Flynn Effect and Its Reversal Are Both Environmentally Caused. *Proceedings of the National Academy of Sciences*. Epub ahead of print.

Burke, Edmund. 1834. Reflections on the Revolution in France. In *The Works of Edmund Burke*. London: Holdsworth and Ball.

Campbell, Jared, Susan Bellman, Matthew Stephenson, and Karolina Lisy. 2017. Metformin Reduces All-Cause Mortality and Diseases of Ageing Independent of Its Effect on Diabetes Control: A Systematic Review and Meta-Analysis. *Ageing Research Reviews*. 40: 31-44.

Chalmers, David. 2010. The Singularity: A Philosophical Analysis. *Journal of Consciousness Studies*. 17(9-10): 7-65.

Ćirković, Milan, and Robert Bradbury. 2006. Galactic Gradients, Postbiological Evolution, and the Apparent Failure of SETI. *New Astronomy*. 11(8): 628-639.

Collings, David. 2014. *Stolen Future, Broken Present: The Human Significance of Climate Change*. Ann Arbor, MI: Open Humanities Press.

Dawkins, Richard. 1995. *River Out of Eden: A Darwinian View of Life*. New York, NY: Basic Books.

de Grey, Aubrey. 2003. An Engineer's Approach to the Development of Real Anti-Aging Medicine. *Science of Aging Knowledge Environment*. 2003(1).

de Grey, Aubrey. 2004. Escape Velocity: Why the Prospect of Extreme Human Life Extension Matters Now. *PLOS*. 2(6).

Delon, Nicolas, and Duncan Purves. 2018. Wild Animal Suffering is Intractable. *Journal of Agricultural and Environmental Ethics*. 31(2): 239-260.

Falk, an. 2018. Why Some Scientists Say Physics Has Gone Off the Rails. NBC. https://www.nbcnews.com/mach/science/why-some-scientists-say-physics-has-gone-rails-ncna879346.

Feynman, Richard. 1965. *The Character of Physical Law*. Cambridge, MA: The MIT Press.

Frick, Johann. 2017. On the Survival of Humanity. *Canadian Journal of Philosophy*. 47(2-3): 344-367.

Greaves, Hilary. 2017. Population Axiology. *Philosophy Compass*. 12(11): 1-21.

Häggström, Olle. 2016. *Here Be Dragons: Science, Technology, and the Future of Humanity*. Oxford: Oxford University Press.

Hanson, Robin. 1998. The Great Filter: Are We Almost Past It? http://mason.gmu.edu/~rhanson/greatfilter.html.

Hanson, Robin. 2016. *The Age of Em: Work, Love, and Life When Robots Rule the Earth*. Oxford: Oxford University Press.

Hughes, James. 2006. The Illusiveness of Immortality. In Charles Tandy (ed.) *Death And Anti-Death, Volume 3: Fifty Years After Einstein, One Hundred Fifty Years After Kierkegaard*. Ann Arbor, MI: Ria University Press.

Jolliffe, Theo. 2015. Has the First Person to Achieve Immortality Already Been Born? *Motherboard*. https://www.vice.com/en_us/article/9akmp5/has-the-first-person-to-achieve-immortality-already-been-born.

Kaku, Michio. 2005. *Parallel Worlds: A Journey Through Creation, Higher Dimensions, and the Future of the Cosmos*. New York, NY: Doubleday.

Kaneda, Toshiko, and Cal Haub. 2018. How Many People Have Lived on Earth? Population Reference Bureau. https://www.prb.org/howmanypeoplehaveeverlivedonearth/.

Knoll, Andrew, and Richard Bambach. 2000. Directionality in the History of Life: Diffusion from the Left Wall or Repeated Scaling of the Right? *Paleobiology*. 26(sp4): 1-14.

Kurzweil, Ray. 2005. *The Singularity is Near: When Humans Transcend Biology*. New York, NY: Viking Penguin.

Locke, John. 1689. *An Essay Concerning Human Understanding*. London: William Tegg & Co., Cheapside.

MacAskill, Will. 2014. *Normative Uncertainty*. Dissertation. http://commonsenseatheism.com/wp-content/uploads/2014/03/MacAskill-Normative-Uncertainty.pdf.

Metzinger, Thomas. 2017. Benevolent Artificial Anti-Natalism (BAAN). *Edge.org*. https://www.edge.org/conversation/thomas_metzinger-benevolent-artificial-anti-natalism-baan.

McGinn, Colin. 1993. *Problems in Philosophy: The Limits of Inquiry*. Oxford: Blackwell Publishers Ltd.

Medvedev, Zhores, and Roy Medvedev. 2006. The Unknown Stalin. New York, NY: I.B. Tauris & Co. Ltd.

Merkle, Ralph. 2018. Information-Theoretic Death. Unpublished manuscript. http://www.merkle.com/definitions/infodeath.html.

Miles, Robert. 2017. What Can AGI Do? I/O and Speed. YouTube. https://www.youtube.com/watch?v=gP4ZNUHdwp8.

Mulgan, Tim. 2009. Rule Consequentialism and Non-Identity. In Melinda Roberts and David Wasserman (eds.), *Harming Future Persons: Ethics, Genetics, and the Nonidentity Problem*. New York, NY: Springer.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Parfit, Derek. 2017. *On What Matters: Volume Three*. Oxford: Oxford University Press.

Paul, L.A. 2014. *Transformative Experience*. Oxford: Oxford University Press.

Pearce, David. 1995. *The Hedonistic Imperative*. https://www.hedweb.com/hedethic/tabcon-hi.htm.

Pearce, David. 2007. Selection Pressure an Radical Anti-Natalism. https://www.abolitionist.com/anti-natalism.html.

Persson, Ingmar, and Julian Savulescu. 2012. *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.

Pigliucci, Massimo. 2014. Uploading: A Philosophical Counter-Analysis. In Russell Black ford and Damien Broderick (eds.) *Intelligence Unbound: The Future of Uploaded and Machine Minds*. New York, NY: Wiley-Blackwell.

Pinker, Steven. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York, NY: Penguin Books.

Ranj, Brandt. 2016. Google's Chief Futurist Ray Kurzweil Thinks We Could Start Living Forever by 2029. *Business Insider*. https://www.businessinsider.com/googles-chief-futurist-thinks-we-could-start-living-forever-by-2029-2016-4.

Russell, Bertrand. 1954. *Human Society in Ethics and Politics*. New York, NY: Routledge.

Sandberg, Anders. 2014. Being Nice to Software Animals and Babies. In Russell Blackford and Damien Broderick (eds.) *Intelligence Unbound: The Future of Uploaded and Machine Minds*. New York, NY: Wiley-Blackwell.

Sandberg, Anders, and Nick Bostrom. 2008. Whole Brain Emulation: A Roadmap. Future of Humanity Institute Technical Report #2008-3. https://www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf.

Sandberg, Anders, Eric Drexler, and Toby Ord. 2018. Dissolving the Fermi Paradox. arXiv. https://arxiv.org/pdf/1806.02404.pdf.

Scharf, Caleb. 2016. Where Do Minds Belong? *Aeon*. https://aeon.co/essays/intelligent-machines-might-want-to-become-biological-again.

Scheffler, Samuel. 2007. Immigration and the Significance of Culture. *Philosophy & Public Affairs*. 35 (2): 93-125.

Scheffler, Samuel. 2016. *Death and the Afterlife*. Oxford: Oxford University Press.

Scheffler, Samuel. 2018. *Why Worry About Future Generations?* Oxford: Oxford University Press.

Schneider, Susan. 2008. Future Minds: Transhumanism, Cognitive Enhancement, and the Nature of Persons. *Neuroethics Publications*. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1037&context=neuroethics_pubs.

Schneider, Susan. 2016. It May Not Feel Like Anything To Be An Alien. *Nautilus*. http://cosmos.nautil.us/feature/72/it-may-not-feel-like-anything-to-be-an-alien.

Schneider, Susan, and Joe Corabi. 2014. The Metaphysics of Uploading. In Russell Blackford (ed.) *Uploaded Minds*. New York, NY: Wiley-Blackwell.

Temkin, Larry. 2008. Is Living Longer Living Better? *Journal of Applied Philosophy*. 25(3): 193-210.

Senthilingam, Meera. 2017. Seven Reasons We're at More Risk than Ever of a Global Pandemic. CNN. https://www.cnn.com/2017/04/03/health/pandemic-risk-virus-bacteria/index.html.

Tomasik, Brian. 2016. Should We Intervene in Nature? Reducing Suffering. http://reducing-suffering.org/should-we-intervene-in-nature/.

Tomasik, Brian. 2017. The Importance of Wild-Animal Suffering. Foundational Research Institute. https://foundational-research.org/the-importance-of-wild-animal-suffering/.

Tonn, Bruce. 2009. Obligations to Future Generations and Acceptable Risks of Human Extinction. *Futures*. 41(7): 427-435.

Turchin, Alexey, and Maxim Chernyakov. 2018. Classification of Approaches to Technological Resurrection. Unpublished manuscript. https://philarchive.org/archive/TURCOA-3.

Verdoux, Philippe. 2011. Emerging Technologies and the Future of Philosophy. *Metaphilosophy*. 42(5): 682-707.

Wade, Nicholas. 2004. Comment on "The Impact of the Revolution in Biomedical Research on Life Expectancy by 2050. In Henry Arron and William Schwartz (eds.) *Coping With Methuselah: The Impact of Molecular Biology on Medicine and Society*. Washington D.C.: Brookings Institution Press.

Walker, Mark. 2002. Prolegomena to Any Future Philosophy. *Journal of Evolution and Technology*. 10. https://www.jetpress.org/volume10/prolegomena.html.

Walker, Mark. 2008. Cognitive Enhancement and the Identity Objection. *Journal of Evolution and Technology*. 18(1): 108-115.

Ward, Peter, and Donald Brownlee. *Rare Earth: Why Complex Life is Uncommon in the Universe*. New York, NY: Copernicus Books.

Wiblin, Robert. 2017. Why the Long-Term Future of Humanity Matters More Than Anything Else, and What We Should Do About It. 80,000 Hours. https://80000hours.org/podcast/episodes/why-the-long-run-future-matters-more-than-anything-else-and-what-we-should-do-about-it/.

Williams, Bernard. 1973. *Problems of the Self: Philosophical Papers, 1956-1972*. Cambridge: Cambridge University Press.

[1] I refer here to situations in which continued existence is unequivocally worse than extinction. I have elsewhere described this as a "hyper-existential risk" catastrophe.
[2] That is, zoophilic because Benatar's contentions apply more broadly to all sentient life. Given the prior meaning of this term as "having an attraction to or preference for animals," I suggest the term "*sentiophilic*" instead.
[3] See also Benatar 1997.
[4] Note that I am revising this manuscript in the early phases of the SARS-CoV-2 outbreak. I do not know, as of this writing, how the situation will unfold, but no doubt most readers will.
[5] A similar point could be made about the misery caused by factory farming and the Darwinian struggle for existence (see Tomasik 2016, 2017). Consider Richard Dawkins' (1995) observation that "the total amount of suffering per year in the natural world is beyond all decent contemplation. During the minute it takes me to compose this sentence, thousands of animals are being eaten alive; others are running for their lives, whimpering with fear; others are being slowly devoured from within by rasping parasites; thousands of all kinds are dying of starvation, thirst, and disease." This relates to footnote 1 above.
[6] Where a "wrong" action is defined as "an action that causes harm."
[7] Furthermore, the addition of "or bad for" clearly applies to Benatar's formulation of anti-natalism, for reasons relating to the harm-benefit asymmetry.
[8] In his words, "my arguments in [chapter 6] and previous ones imply that it would be better if humans (and other species) became extinct" and, elsewhere, "extinction … would result from universal acceptance of my view" (Benatar 2006).
[9] See also Kahane 2014.
[10] Bostrom continues: "The idea here is that you have some evaluation standard that is fixed, and you form some overall plan to achieve some high-level subgoal. This is your idea of how to maximize this evaluation standard. A crucial consideration, then, would be a consideration that radically changes the expected value of achieving this subgoal, and we will see some examples of this." (Bostrom 2014a).
[11] But of course new forms of life could emerge in the new universe that this would create.
[12] Of course, some humans could naturally live longer than 80 years—up to at least 122 years—but for the present purposes, we can bracket this.
[13] I would like to thank an insightful anonymous reviewer for encouraging me to develop this idea.
[14] In brief, general problems include the question of how society will need to be restructured to accommodate people who will never retire and whether people with indefinitely long lives will suffer from crushingly oppressive ennui—especially if, Soren Kierkegaard (1852) speculated, "boredom is the root of all evil." There are also concerns about overpopulation, associated today with climate change, ecological collapse, and other environmental ills, if humanity continues to procreate without older generations dying off (see Kuhlemann 2018). Yet ceasing to procreate could remove a major source of value for many people; as Larry Temkin (2008) writes, "if the cost of immortality would be a world without infants and children, without regeneration and rejuvenation, it wouldn't be worth it." Even more troublesome is the question of who gets to be included and excluded in the final generation? What happens if apocalyptic terrorists, deranged dictators, genocidal madmen, violent psychopaths, and dangerous lone wolves gain access to technologies that could essentially confer eternal life? Consider that, at the behest of Joseph Stalin, "life-extension became a central subject of Soviet medical research" (Medvedev and Aleksandrovich 2006). My decision not to dwell on such issues is purely a matter of space limitations. Let's now turn to the issues above:
[15] That is, on a realist interpretation of personhood. Pigliucci (2014) himself argues that "there is … no *metaphysical fact* of the matter about personal identity."
[16] Note that if we someday upload, duplicate, or simulate a *large number of minds like ours*, then this would constitute evidence that posthuman civilizations tend to simulate a large number of minds, and this would constitute evidence that we are probably inside a simulation (Bostrom 2003b). Yet another issue worth considering.

[17] It is also worth adding that attempts to emulate entire nervous systems could themselves pose some serious ethical hazards. First, as Bostrom notes, "before we would get things to work perfectly, we would probably get things to work imperfectly" (Bostrom 2014b). The result could be that imperfectly simulated brains experience moments of truly intense suffering, perhaps in the form of psychotic hallucinations or delusions, grand mal seizures, and so on, before a normal state of consciousness and mentality is established. Second, since emulating parts of the human brain will likely antedate emulating a whole human brain, some form of "neuromorphic AI"—that is, a system that combines simulations of human brain regions with synthetic AI patches—will probably occur before whole-brain emulation. Insofar as this AI system is (i) superintelligent, and (ii) has a value system that is not sufficiently aligned with our "human values," then we have reason to worry about it bringing about a killing-extinction event (that is, for reasons pertaining to the "instrumental convergence thesis"). In other words, the road to mind-uploading could itself pose grave threats to our survival (see Bostrom 2014b).

[18] My own intuitions, as it happens, align best with the "no-self" position defended by James Hughes (2006). Indeed, Hughes imagines that mind-uploading will usher in a "post-individual age" whereby "we will be able to copy, share and sell our memories, beliefs, skills and experiences. We will selectively adopt personalities for specific purposes—Machiavelli for politics, Cyrano for love. … Some people will live broadcast VoyeurLives, just as some now put VoyeurCams in their homes, and others will choose to spend a lot of time in someone else's life—like climbing into John Malkovich's head for weeks instead of 15 minutes at a time. … Personalities will begin to bleed and blur. We will write copious reams about the decline of the old discrete, continuous self, and the rise of the new creative, collaborative self-process. … The most dramatic challenges to our social and philosophic world will probably come from hive minds and distributed selves (Hughes 2006)."