On terraforming, wild-animal suffering and the far future

Introduction

Many wild animals are believed to have lives that are not worth living (Tomasik 2015a). That is to say, for the same reasons one might argue a human child is 'better off' not being born if they are to be brought into a life of hardship and suffering, it is argued that many wild animals are better off not having been born. One might look at the life of a large mammal such as a wild lion that enjoys long naps in the sun and wonder how they might have a bad life. Our intuition is thus that nature is a peaceful place. This is because the reproductive strategy of large and obvious animals such as lions is K-selected, which involves giving birth to relatively few young and spending a large amount of time caring for them. This is compared to r-selected reproductive strategies, which involve giving birth to a relatively large number of offspring and spending little time caring for each one. As a result, many die before reaching maturity, often in painful ways. Animals and other beings (including fish, insects and other invertebrates) of this type are significantly more numerous in the wild.

Even for those wild animals that don't die young, the rest live in a high state of stress; avoiding predators, being constantly hungry and thirsty, and suffering from the elements, disease and accidental injury. If we can conclude that the vast majority of wild animals (from this point, the term wild animals also includes insects) have lives that are, on balance, bad, then we could also conclude that, all else being equal, increasing the number of wild animals in existence would be bad, as it would increase the total amount of suffering. If humanity spread wild animals to other planets in the future, this could be astronomically bad.

This possibility is complicated by the likelihood of an artificial general intelligence experiencing an intelligence explosion, and therefore being in control of the future of sentient minds. It has been argued that this will likely occur long before terraforming of other planets (e.g. altering the atmosphere to make them habitable for humans) or mass space colonisation becomes feasible (e.g. Tomasik 2016a), though for reasons I will discuss in this essay, I don't believe this is a foregone conclusion.

In this essay, I discuss some of the technical and ethical considerations associated with terraforming and colonising the universe, and the risk of spreading wild-animal suffering (or some other bad outcome) to other planets. I also discuss the likelihood of spreading wild-animal suffering, and propose some ways that it could be made less likely or 'less bad' through actions today, e.g. by values spreading. I then highlight some open research questions which, based on my research, I propose are high value to work on. An examination of the various ethical codes in relation to this question is crucial, as many arrive at different (and in some cases, opposite) solutions to the problem of wild-animal suffering, or suffering in general. In particular, negative utilitarianism and classical hedonistic utilitarianism appear to have a potential conflict.

Due to the nature of an interdisciplinary body of research, this essay has not fully covered the literature and considerations in each field discussed. Further, I do not place 100% confidence on the ideas and solutions I propose. However, the essay is intended to provide a useful starting point for those interested in the question of how to minimise the risk of spreading suffering through the universe (and how to maximise good outcomes), to highlight some useful areas for further research, attempt to make some original contributions to the question at hand, and to encourage dialogue on high impact topics.

Why is wild-animal suffering worthy of moral consideration?

Some might ask why wild-animal suffering is worth doing anything about it. The shorthand argument for valuing the wellbeing of animals is that they can suffer and experience wellbeing just as humans can, and so if we value wellbeing generally (that is, we take some utilitarian view on ethics), we should value it for animals too. However, even those who agree that the wild is full of suffering will often argue that nature should be left alone. Others, particularly in the animal rights/abolitionist movement, argue that eliminating factory farming is important because humans caused it, but eliminating wild-animal suffering is not an obligation that humans have, since they had no direct influence on the suffering of some animal being eaten alive by parasites. If we take an ethical stance with some form of utilitarianism, we should try and reduce suffering and maximise wellbeing where possible, regardless of where it is and how it is caused.

Applying the concept of normative uncertainty (MacAskill 2014) we can see a justification for working to reduce wild-animal suffering even if we're not certain that we are morally obliged to do so. Let us suppose we think that there is some chance that we are morally obligated to reduce wild-animal suffering, and some chance that we are not. In the case where we are morally obliged to try and reduce wild-animal suffering (or at least consider it), the morally best action is to reduce it. If we are not obliged to reduce wild animal suffering, reducing wild-animal suffering and ignoring it would be equally as 'choice-worthy'. Therefore in this case, reducing wild-animal suffering seems like the better choice.

Of course, this is not a completely satisfying answer, as some might argue (as abolitionists or rights-based animal advocates do) that animals should be not exploited. If there is some chance that we are morally obliged to *not* exploit animals (even if it means reducing suffering), this complicates the question of what we ought to do. Thus, for a complete answer using normative uncertainty, we must consider the probabilities of each of these moral theories being correct. This is beyond the scope of this essay, though Beckstead (2013) explores reasons for a form of utilitarianism being more likely to be a true moral theory than a form of deontology.

Tomasik (2015b) has raised the point that theologians often ask why a good god would create so much suffering in the world, both in the wild and in human society. We are in an opportunity to ask this question of ourselves now. If we could control the future, and we see ourselves as being good, why would we choose to create a universe full of suffering over one full of wellbeing, or at least an absence of suffering? By not acting on wild-animal suffering, we are choosing a world full of it.

Why is the far future worthy of moral consideration?

Beckstead (2013) argues that, because humanity and its descendants may survive for billions of years or more, and that actions taken today can meaningfully affect the trajectory of humanity, what matters most today may well be the effects that we have on the far future. This can be achieved by altering the rate of human progress, through reducing the risk of human extinction, or through changing the trajectory of humanity, e.g. spreading better values to the next generation could have persistent effects. As will be discussed in a later section, Bostrom (2003) argues that any delay in progress towards colonising a significant portion of the universe and filling it with conscious minds experiencing wellbeing is astronomically wasteful.

One possible reason for not caring about the far future might be to say that beings who don't exist yet aren't worthy of moral consideration. I discuss this at length in a later section, but I can offer one brief thought experiment for why I believe most people don't actually think this way. Imagine a childcare centre with twenty children aged 6 present at any one time. To simplify, there is never an adult present. If this childcare centre were blown up tomorrow, it is clear that this is a bad thing. Let's now suppose though that someone set up a timed bomb to explode in ten years. Once set, there is nothing that can be done about the bomb. The children who will be killed by this bomb have not yet been conceived, but most people would generally agree that the act of setting up this bomb is still a bad thing. Therefore, most people should agree that any act performed today which makes the future worse is bad.

The complication of digital sentience and an intelligence explosion

As has been discussed by a number of researchers (e.g. Tomasik 2015b), it is not clear that terraforming and space colonisation, if ever performed at all, will be performed under the control of humanity or its descendants. It is feasible that an artificial general intelligence (AGI) will reach singularity (rapidly become vastly more intelligent and capable than humans) and gain total control over humanity's future (or end humanity) long before terraforming becomes feasible. Based on a survey of experts (Müller & Bostrom 2016), the arrival of a high-level machine intelligence predicted to have a 1 in 2 chance of occurring by 2040-2050, and a 9 in 10 chance by 2075. Such a system is then predicted to result in AGI in less than another 30 years. Terraforming, however, is not considered likely to be partially complete for another 100 years (discussed in more detail below in 'Terraforming Mars').

Such an artificial intelligence may not share the goals of humanity today or take the most utilitarian actions. It is estimated that there is a one in three chance that the development of superintelligence will be bad for humanity (Müller & Bostrom 2016).

Digital sentience as created by a hypothetical artificial intelligence could potentially be considered to have some level of sentience and thus moral consideration. It is not guaranteed that such sentience would be well cared for by some overarching AI, or whether they would have good lives. If digital sentience has, on average, net negative lives, spreading digital sentience throughout the universe could be bad.

There is a substantial literature on AI safety, risks and digital sentience that this essay assumes some familiarity with. For a complete overview of the technical and strategic considerations, see *Superintelligence* (Bostrom 2014).

Terraforming Mars

The terraforming literature appears to focus primarily on Mars, and for a variety of technical reasons discussed below it seems reasonable to assume that Mars will be the first off-Earth body to be terraformed. Thus, Mars is the key focus in this essay. Fogg (2011), McKay et al (1991), Marinova et al (2005) and Graham (2004) overview some of the technical considerations associated with terraforming Mars. Fogg (2011) suggests that Mars may be capable of supporting animal life in the future, but argues that this won't happen any time soon.

Zubrin (2011) suggests that Mars has all of the requirements to be capable of supporting life (e.g. gravity, rotation rate, elements) except a sufficient atmosphere, which is primarily composed of carbon dioxide and is around 1% the pressure of Earth's. There is substantial evidence that its atmosphere was once thicker (Zubrin 2011).

Zubrin (2011) proposes that there are sufficient reserves of carbon dioxide on Mars in the ice caps and in the regolith to thicken the atmosphere if it is released. Zubrin estimates that releasing all of the carbon dioxide reserves would result in an atmosphere 30% that of Earth's, which can be achieved via global warming. This might be possible via a sustained warming at the south pole of 4 degrees C, which would result in a runaway warming effect as more carbon dioxide is released. For a full technical discussion of this process, see Zubrin (2011).

This process remains subject to much uncertainty, particularly in modelling how rapidly the CO_2 locked in the regolith will be released and how much CO_2 is present. Zubrin (2011) claims that, given some assumptions about regolith properties, one would have to outgas the CO_2 to a depth of 200 meters to produce air pressures equal to that of Earth at sea level. Zubrin estimates that it would take approximately 2,500 years to fully outgas to that depth. It is also worth noting that this assumes that regolith exists to that depth at all points on the planet, which is an unreasonable assumption given that solid rock is known to outcrop in some places on the surface of Mars, and thus I would estimate it would take even longer than this.

Zubrin (2011) proposes several methods of inducing the initial warming required to outgas the CO_2 from the regolith and melt the ice at the poles, including:

- Orbital mirrors to heat the polar ice,
- Mass production and release of artificial halocarbon (CFC) gases,
- Release and generation of bacterial ecosystems,

Or some combination of the three. A full discussion of the technical feasibility of these techniques is beyond the scope of this essay, however a few comments will be made below.

Of these solutions, the mass production of CFC appears to require the lowest levels of technological capability, but it does require the construction of a powerplant with around 5,000 MWe output (around that required for a large city in USA) and a budget of several hundred billion US dollars. This may be optimistic, as CFCs on Mars would be substantially less long lived in the Martian atmosphere than on Earth due to the lack of UV shielding and an ozone layer (McKay et al 1991).

Zubrin (2011) suggests that, using these techniques, the temperature on Mars could be made acceptable for human conditions without space suits within ‘decades’ (assumed here to refer to 50 years) of initiation, although an oxygen supply would still be required. At this point, some types of plants may be capable of spreading across Mars (Zubrin 2011). It is expected that, if plants are able to spread across the surface, some insects may be able to as well, whether intentionally or by accident. Therefore, we may be less than a century from irreversibly (at least in the mid-term) spreading wild-animal (insect) suffering to Mars. This is likely a lower bound for how long it will take, as proponents of Mars colonisation have been saying it is just a few decades away for a few decades (e.g. Zubrin 1991).

Zubrin (2011) estimates that Mars would be livable for humans and other animals without habitats, space suits or breathing masks around 900 years after the terraforming process begins. McKay (2007) has suggested that Mars could be sufficiently warmed for human livability in 100 years (from commencement), while generating an oxygen rich atmosphere would take the significantly longer time of 100,000 years or more.

In order to provide a more accurate assessment of the feasibility and timeline of Martian terraforming, further research is required. In particular, the following lines of research appear to be the major blockages:

- Studying the dynamics of regolith outgassing,
- Estimating or measuring the amount and distribution of CO_2 and water etc in the Mars subsurface,
- Adapting Earth climate models to predict the impact of terraforming Mars,
- Further work to determine the economic, technical and social/political feasibility of proposed terraforming technology.

McKay (2007) suggests that the loss of atmosphere due to solar wind and other factors played a major role in its loss of atmosphere. These effects are more pronounced on Mars than Earth because of its smaller size. McKay (2007) therefore suggests that maintaining a terraformed atmosphere may not be feasible on a time scale of millions of years. Brown et al (2015) found that the current rate of atmosphere loss due to solar wind is around 100 grams per second, which might be used to estimate the sustainability of terraforming efforts.

Ethics of terraforming

Many argue that, if Mars were to be discovered to have pre-existing microbial life of some type, that it would be unethical to terraform it (e.g. McKay 1990). From a strictly utilitarian point of view, there is no obvious reason as to why this is the case. Past discussions of the ethics of terraforming appear to rarely include any consideration of utilitarianism. McKay (2007) and Sparrow (2015) appear to be the only exceptions. McKay suggests that 'Wise stewardship' is one system of environmental ethics (that "*The fundamental principle that the measure of all things is utility to humans, in the broadest and wisest sense of utility.*"), though this definition does not include non-human animals or the possibility of digital sentience.

It is interesting to note that even Robert Zubrin, arguably one of the most avid proponents for the terraforming of Mars, does not use a completely utilitarian argument for doing so. Zubrin (2011) states that "*I would say that failure to terraform Mars constitutes failure to live up to human nature and a betrayal of our responsibility as members of the community of life itself.*"

McKay (2007) suggests that some potential utilitarian reasons for terraforming Mars include using it as a backup colony (though Matheny (2007) argues that building refuges on Earth would be more efficient). McKay is critical of the technical feasibility of this, noting that even if most people on Earth die from a catastrophic event, the population would still likely be larger than on Mars. However, McKay neglects the possibility of an existential-risk event occurring on Earth which may result in the complete extinction of humanity, for example, global thermonuclear warfare, an intelligence explosion (AGI), or bioterrorism. Of course, many of the most likely existential risk events (e.g. AGI) are expected to not just affect Earth, and so even Mars wouldn't be safe. Some researchers have proposed that the probability of human extinction by 2100 is as high as 1 in 2 (e.g. Rees 2003).

Another reason for terraforming would be to derive value from the use of Mars today. The example McKay (2007) gives is of using Mars to advance scientific understanding and improve life on Earth. McKay does not appear to take seriously, or at least doesn't discuss, the possibility of using Mars as an extra planet to spread and increase the number of humans or animals living lives worth living.

Schwartz (2013) discusses several of the prevailing views on the ethics of terraforming and colonising space using environmental ethics. It is noted that a common objection to terraforming space is "*Why expend so much energy studying space, when there are so many problems to solve here on Earth?*". Which leads to tension between space exploration advocates and environmentalists.

Schwartz (2013) says that if there is life on some planet, a non-anthropocentrist (believing that the nature has intrinsic value) would say that the choice would be clearly to not terraform it. If there is no life, an anthropocentrist (believing that nature's value depends on the value humans derive from it) would say that the choice is clearly to terraform it for the benefit of humanity. Schwartz suggests the answer would be less clear for each group in the opposite case.

Schwartz (2013) also notes that exploring other planets has benefits for understanding Earth's own ecosystem. For example, understanding Venus helped us understand the greenhouse gas effect, which was useful for understanding climate change on Earth. Schwartz (2013) states that their initial assessment is that it would be 'morally recommended' to terraform if it provided useful environmental knowledge, assuming humans are morally obligated to gain environmental knowledge, and morally neutral if humans were not obligated to do so.

If a planet is discovered to have no life, by many standards there would be no cause for immediate concern. If a planet is discovered to have microbial life that one day may become intelligent life, I would argue that there is still no cause for concern. If what we value is wellbeing, and we suspect we can reasonably hope to fill a planet with minds experiencing wellbeing, this would be far more efficient than waiting for several million years of evolution or more to run its course, complete with all of the suffering associated with evolution in the hopes that sentient life will arise, without any guarantee that it will share our values in the future.

The Curiosity probe which was sent to Mars was found to have 65 separate bacterial species despite decontamination efforts, and it is possible that some of these survived the trip and entry to Mars (Madhusoodanan 2014). Therefore, it may already be too late to worry about keeping Mars uncontaminated. The UN subsequently implemented a treaty to stop other planetary bodies from being contaminated, lest a future space mission to detect life mistakes Earth life for extraterrestrial life.

A potential terraforming/WAS spreading scenario

This section outlines a possible scenario for the colonisation of other planets and star systems in a universe where there is no concern for wild-animal suffering, and where the control problem (and other problems) for artificial intelligence have been solved, and thus AGI works for the benefit of humanity within the bounds of strict instructions. This scenario is loose and is meant only to be illustrative of a potential outcome, and the order that some events occur in may differ. A number of other assumptions about the values of humanity in the future are made, for example that resistance to colonising and terraforming Mars are overcome, which includes overturning the part of the Outer Space Treaty (United Nations 2002) that refers to contamination.

1) The first human colony on Mars is established inside shelters in 2045.

Shishko et al (2015) suggest that humans will first land on Mars in the 2030-2040 range, a Mars colony will be established in the 2040-2050 range and some form of economic pipeline between Earth and Mars will be underway from 2050.

2) Mars is partially terraformed by 2155-4605.

The technological and economic feasibility and political will to begin the terraforming process on Mars is achieved 50 years after the first colony is established (in the year 2105, which is an educated guess). It seems reasonable to believe Mars will be mostly habitable for humans (e.g. only requiring a breathing mask) 50 to 2,500 years from commencement of terraforming (2155 to 4605). This is a large range, but it highlights that it could potentially happen relatively soon.

3) Mars is seeded with wild animals for aesthetic and nostalgic reasons (or accidentally) at some point after 2155.

McKay (2007) claims that the air on Mars wouldn't be breathable for 100,000 years, but this seems unreasonably long given the rate of advance of human technology. However, even if this is the case, it is plausible that there are some insect species that can survive in a partially terraformed atmosphere, and so it is possible that one of these species will spread across Mars by accident, or for some proposed ecosystem/terraforming development related reason.

4) The remainder of the Solar system is colonised, with manned structures on asteroids and the moons of other planets. Some bodies, where feasible, are terraformed and seeded with wild animals.

The timescale for the terraforming of other bodies is expected to be significantly longer than for Mars, however the process can be commenced long before Mars terraforming is completed (and it is expected that this will be the case; if humanity decides to terraform one world, it is unlikely that we would stop there). Colonies in enclosed structures on other bodies may be created prior to the terraforming process beginning on Mars.

5) The vast resources of the Solar system are increasingly utilised.

Examples include creating a Dyson Sphere (Dyson 1960) to harness the majority of the solar power from the Sun, asteroids are harvested and/or turned into space colonies, and planets themselves are increasingly stripped and turned into space habitats. The timescale on this is hard to predict, though it seems unreasonable to believe that it would take any less than 10,000 years given the current rate of technological advance (e.g. see Hanson 2009), unless there is a major shift towards lower or negative growth, or there is a catastrophic event.

Space habitats may be more numerous and abundant than terraformed planets in the long run of the universe, given their additional capacity for habitable space for life (C. Shulman pers. Comms. 2016). About 30% of sunlight reaching Earth bounces off of the clouds, atmosphere and land. Around 28% of the remaining 70% falls on land (Murphy 2011), and much of this land is not biologically productive (e.g. deserts). Thus, it is reasonable to assume that planets will be converted to structures that can more efficiently use the energy of stars (e.g. Dyson Spheres).

It is not completely clear how the use of space habitats instead of terraformed planets will affect the existence of wild animals. The purpose built nature of space habitats for humans may mean that wild animals are not included, though given the tendency of humans today towards conservation of individual species, it seems plausible that wild animals will be retained, e.g. in some form of space zoo. If this is the case, wild-animal suffering may be substantially small (and not quite wild anymore). However, it is also possible that some planets or parts of planets are maintained as nature preserves with wild animals intact.

6) Interstellar colony ships are launched.

7) Colonisation technology improves with each colony ship iteration, leading to faster travel and faster colonisation and relaunch at each planetary body.

8) Intergalactic colony ships are launched.

9) Colonisation speed eventually approaches the speed of light once physical limits are approached (Hanson 1998).

It will be unlikely for interstellar colony ships to be piloted by humans. Von Neumann Probes are theorised self-replicating probes which spread by colonising planets, rebuilding themselves based on a set of programmed instructions, and relaunching indefinitely (Von Neumann & Burks 1966). Such probes may be designed to spread biological or digital life throughout the universe.

Hanson (1998) proposes a decision making model for how a civilisation might act if it were colonising outwards from an origin point in a wave-like manner. This has useful properties for predicting what the far future might look like if it remains under human control.

Pearce (2004) has suggested the possibility of 'cosmic rescue missions' whereby we are able to find and reduce suffering already existing on other planets, though Tomasik (2016a) is skeptical about this concept, as it's not obvious there are lots of sentient beings waiting to be rescued. It's also unclear whether environmentalists (or humanity in general) would support such missions. But perhaps there are extraterrestrials running WAS simulations whom we could convince to stop (persuasively or aggressively).

This is not to say that the future will necessarily take this path, or even a path like it, but we can use it as an illustrative example of what may occur as the result of space colonisation. As stated, this scenario simplifies the complication of AGI for simplicity. An AGI not sufficiently controlled or loaded with the values of humanity would be expected to not care about terraforming planets or spreading wildlife. However, it may be inclined to simulate wildlife, which could also be bad (see below). It is not initially clear why an AGI would do this, unless it was pre-loaded with some value to do so. If the values loading problem was sufficiently overcome by this time, presumably the AGI would work towards optimising conscious wellbeing over the remaining course of the universe, and thus any further considerations become null.

How many animals (or insects) could Mars support?

It is concerning that, after only a brief search, a number of references to the 'promise' of using insects as a food source on Mars or as part of the terraforming process could be found (e.g. Yamashita et al 2009). The current prevailing sentiment is certainly that spreading insects to Mars would be good, or at least morally neutral.

There may be insects on Earth which are capable of surviving on the surface of a partially terraformed Mars (or insects might be genetically modified to survive). Such insects might be spread by accident, as a food source, or as a mechanism to terraform Mars. It is worth estimating how many insects a partially or fully terraformed Mars might be capable of supporting to get a sense of how bad it would be to spread insects to Mars in the long run.

Tomasik (2016b) has suggested that the number of insects on Earth is 10^{17} to 10^{19} (not including other creatures such as nematodes). The surface area of Mars is approximately 83% the land surface area of Earth. While land on Mars, even after terraforming, might be expected to be less capable of supporting life than land on Mars (e.g. due to lack of organic matter), we could use this as an upper bound. The vast majority of sentient beings on Earth's land are insects, and while we might give insects less moral consideration than large animals due to there being some uncertainty about their sentience, and a likelihood of being less sentient if they are, it is still reasonable to assume that insect suffering on Earth dominates (see Tomasik (2016c) for a summary of literature on insect sentience). Imagine creating a new world with up to 83% the amount of suffering on Earth's land (not including ocean suffering), and very little of the wellbeing. This is what is at stake.

Dickens (2016a) estimates that, if wild animals were spread throughout the universe, there would be $3x10^{39}$ to $4x10^{44}$ (80% confidence interval) wild animals (with insects/bugs discounted according to their probability of being sentient and their level of sentience if they are). This estimation is highly simplified, and I would encourage a more rigorous estimation to be undertaken to determine what is at stake on an astronomical scale over the course of the universe.

If all life is negative

Benetar (2008) argues that coming into existence is always a harm, as all lives contain some suffering. To not bring one into existence is to eliminate this suffering from occurring. Benetar also claims that, since there is no person in existence before they are born, there is no way to deny someone of wellbeing by not bringing them into existence. This is known as a 'person-affecting view', and is not compatible with classical hedonistic utilitarianism (which Benetar admits). Further, Beckstead (2013) spends much time discrediting person-affecting views, and so I won't spend too long on it here. However, I do want to add a few points to the discussion and examine some possible implications.

Sam Harris proposes a hypothetical worst possible world (WPW) which is full of as much suffering as possible for as many beings as possible (Harris 2011). This is bad, if the word bad means anything at all. It follows that moving away from this is good, with the best possible world (BPW), which is full of wellbeing and void of suffering, on the other side. Benetar's view implies that being in the BPW is equally as good as no existence at all. This does not make sense from the point of view of a classical utilitarian. Even a near-BPW with a fraction of a second of suffering for each conscious mind might be expected to be, on balance, good. Does a second of pain make the rest of the life of enjoyment not worth it? A classical utilitarian would certainly think otherwise.

If we accept this extreme case, we're still left with the reality that we are not in a near-BPW yet, and probably won't be for some time (if ever). Most people in a western society believe their life to be worth living, though Benetar (2008) suggests that this is the result of evolutionary bias (e.g. those with a pro-natal bias are more likely to pass on their genes than those without, and so this trait propagates). This is a valid point, and I do not have a satisfactory way of estimating the strength of this effect and how much it might lead one to believe their life is better than it is. Determining this should be a priority. If the pro-natal bias is sufficiently strong, it is possible that humans always have a net negative life and are incapable of having a net positive life.

If Benetar (2008) is right, and even the BPW is worse than (or as good as) total non-existence, increasing the probability of an existential risk event would become a top priority. This is surely an uncomfortable outcome, and many would be biased towards believing (or hoping) that this is not the case. If this did turn out to be true, the question becomes what to do about it. In what meaningful way could one or more individuals increase existential risk? Such a stance could surely not be made public, as it is even more ridiculous to the public and outside the Overton Window than reducing existential risk. Even without taking a person-affecting view, one might think that the expected suffering in the future outweighs the expected wellbeing, and that, in terms of expected value, increasing existential risk is the best thing to do.

I stress that I do not believe that this is the case or that existential risk should be increased, however the possibility should be concerning, as it would imply that even spreading sentience throughout the universe in the best way possible would be bad (or at least neutral). It is thus worth considering for the purpose of possible future space colonisation and values loading for AGI.

What to do today about wild-animal suffering and the far future

Dickens (2016a) seeks to quantify the value to the far future of interventions performed today. He examines the effect of having additional AI safety researchers on AI safety, the value of animal advocacy on factory farming, wild-animal suffering, and the treatment of sentient digital systems in the future, and the effect of one type of values spreading (spreading concern for non-human animals to AI researchers). While the models used to estimate these effects are simplistic, there do not appear to be any other serious attempts to do so. It might be reasonable to propose that trying to convince terraforming researchers to care about spreading wellbeing (and not suffering) on planets and throughout the universe. See Dickens (2015) for further work on values spreading.

Tomasik (2016d) suggests that promoting concern for wild-animal suffering (which is what the organisation ‘Animal Ethics’ is doing) is one of the most effective things to do today to reduce suffering in the long run. A part of this may be to seek out allies in the broader animal advocacy movement. Tomasik (2016d) also advises caution, in that there is the possibility of advocating concern for wild animals before the general public is ready.

As the majority of traditional literature on space ethics, particularly that of colonisation and terraforming, takes a non-utilitarian perspective, it may be valuable to promote utilitarianism within the space research community. Caution must be taken to ensure that this results in caution due to the possibility of spreading suffering, rather than resulting solely in increased interest from the potential use of Mars for humanity.

I suspect that spreading the values of concern for wild animals or all life forms in general is one of the best ways to have a positive impact on the far future besides reducing existential risk. However, I suggest that investing in discovering, de-risking and refining interventions of reducing wild-animal suffering today is currently undervalued. Many dismiss working towards reducing wild-animal suffering as they claim it is too hard to do so without possibility of dangerous repercussions. These people ignore the fact that many things humans do already influence wild-animal suffering positively or negatively. The least we can do is surely try to minimise the levels of suffering from actions we already undertake. For example, Tomasik (2016e) has suggested converting grass lawns to gravel to reduce plant biomass and therefore insect suffering. Making concrete advances in the field of reducing wild-animal suffering may change perceptions about how feasible reducing wild-animal suffering is, and therefore shift public perspectives about influencing suffering in the far future as well. Dickens (2016b) has also suggested this, and that arguments for wild-animal suffering being intractable are flawed.

A substantial body of good research and interesting hypotheses on wild-animal suffering, the far future and related topics appear to be in the form of personal blog posts or other online discussions. While this may make the content more accessible to the general public (it eliminates paywalls, for example), I might argue that the traditional academic body at large is largely unaware of these ideas because they don’t appear in the traditional academic literature. As a result, when policy analysts do their research for government decision makers, they get (arguably) an unrepresentative/biased account of how things ought to be. This is glaringly obvious after sampling the traditional academic literature on terraforming, which contains very little if any utilitarian account or serious consideration of the far future.

From this, I would argue that the work of blog posts on these topics should be synthesised and collected into academic and mainstream publications such as academic journals and scientific magazines. I would strongly encourage any researcher working on these questions with a substantial volume of good quality blog posts to summarise their work and submit it in the form of academic papers. Further, it is frowned upon or simply unacceptable to reference blog posts in a government report or academic paper. This may be an argument for creating one or more online, themed journals around Effective Altruism related ideas.

Can this problem wait until the future?

Can we hand off to future generations thinking seriously about trying to solve the problems relating to terraforming and spreading wild-animal suffering? This requires looking at two things; how long do we expect to have before space colonisation, and how likely is it that some existential risk (especially AGI related) event will take place before then.

Most artificial intelligence scenarios include propagation throughout the universe as the end result. If artificial intelligence was to propagate throughout the universe with the values held by most humans today, that is, that wild animals ought to be left untouched or even spread, this would mean that we have to find a solution prior to an intelligence explosion (to load the best values into the AGI). As stated above, there is a 9 in 10 chance of high-level machine intelligence being developed by 2075, with AGI not far behind. Even if there is no intelligence explosion, terraforming is arguably feasible by 2155. In the non-AGI scenario, while we would have substantial time before interstellar space colonisation occurs, it seems reasonable to believe that at least insect suffering could be spread to Mars, and thus we should work on understanding the implications and on spreading good values as soon as possible. In an AGI scenario, spreading good values about wild animals into society and thus loading them into an AGI is also expected to be beneficial for how an AGI acts around creating simulations of nature.

Should we speed up or slow down technological progress?

Assuming life (biological or digital) survives the myriad of existential risk events, there is some probability that the far future will be good, and some probability that it would be bad. If it is more likely to be good than bad, we should (simplistically speaking) speed up technological progress to take full advantage of turning the resources in the universe into wellbeing before the heat death of the universe, and seek to reduce existential risk. If it is more likely to be bad than good, we have several choices. We could take the (arguably most depressing) option of increasing existential risk, if we think that the lack of existence of life is better than pure suffering spread throughout the universe. Alternatively, we could attempt to slow down technological progress until we have improved the odds of the future being good, for example by determining the best values to spread and spreading them.

Bostrom (2003) suggests that any delay in advancing the progress of humanity is bad, because it delays unimaginable future wellbeing (which he terms ‘astronomical waste’). Bostrom loosely estimates that delaying the colonsiation of the Virgo Supercluster by one second is equivalent to about 10^{29} potential human lives lost (and presumably many more non-human lives). However, it may also delay a large amount of future suffering. There is arguably more suffering than wellbeing on Earth today if we include non-human animals in farmed conditions and wildlife (Tomasik 2015a). If we create many more planets with similar conditions to Earth, the total amount of suffering in the Virgo Supercluster from non-human animals may outweigh the wellbeing of humans. Is it more likely than not that future wellbeing will dominate than future suffering if we delay progress? To get a basic estimate, it may be useful to survey field experts.

Bostrom (2003) talks of optimising ‘safety’, i.e. the probability that colonisation will occur, but there is still no guarantee that colonisation will be good. But if we ignore that possibility, there is certainly a huge cost to delaying colonisation.

Bostrom (2013) refers to maxipok, or, “*maximise the probability of an ‘ok outcome’, where an ok outcome is any outcome that avoids existential catastrophe.*” I argue this is too simplified to the point of being counterproductive, as the future could quite possibly contain more suffering than wellbeing (or put another way, expected suffering may outweigh expected wellbeing). Maxipok might be better achieved by increasing existential risk, as a lack of both suffering and wellbeing might be neutral from a utilitarian perspective.

Bostrom (2003) only talks about humans or human descendents, but not non-human animals. He says it’s very likely humans will have a good quality of life if the supercluster were colonised, but the wildlife simulations and wild animals on terraformed planets might not be so lucky.

One might ask whether we should seek to slow (or at least not accelerate) progress until we've sufficiently spread good values. How much spreading would be sufficient? Xie et al (2011) demonstrated through modelling that, if an idea is strongly held by 10% of a society, it will rapidly spread. I suspect that this is highly simplified, and not a one-directional path (e.g. many in some societies used to think that having human slaves was morally acceptable). However values spreading researchers might benefit from cross-disciplinary work of this nature and with other principles from psychology and movement building (e.g. McAdam 1986). We still also need to ensure (or at least be sufficiently certain that) we have picked the right values to spread.

It is also worth noting that technological progress improves our ability to deal with some problems, but it brings about many other problems sooner. For a toy example, advanced space technology has given us some hope that we could move an asteroid out of an Earth-intersecting path, but it may also give some the technology to move an asteroid *into* such a path.

Tomasik (2016a) estimates a ~70% probability that a human initiated colonisation of space would result in more suffering than it reduces. Tomasik is not accounting for added wellbeing and takes a negative utilitarian stance, i.e. suffering is bad, and wellbeing is neutral. To suggest why this might be unreasonable, I would ask anyone who holds this view whether they would prefer oblivion (which I assume is neutral, as in a lack of both suffering and wellbeing), or epic party times (defined as the most wellbeing possible for an individual, devoid of all suffering and with no risk of getting bored). It is possible that I am missing the point, as a negative utilitarian may well just respond by saying that they would prefer oblivion.

Just as Beckstead (2013) claims that person-affecting views are arbitrarily asymmetrical, I suggest that negative utilitarianism is also arbitrarily asymmetrical. It should be noted of course that I am alive and therefore subject to the pro-natal bias (and would be very upset if nothing were better than the BPW), but I can't imagine epic party times being worse than (or as good as) nothing.

Tomasik (2016a) and Knutsson (2016) argue that those interested in reducing suffering, even those that think the future is more likely bad than good (or that life will always be net negative), should still cooperate with (or at least not work against) those interested in reducing existential risk. This is because there are benefits to compromise, and it avoids the risk of a negative stigma being associated with reducing suffering and considering the far future. Thus, suffering reducers should focus on making the future less bad. There are a lot of other strategic considerations for thinking about how to manage progress, and what to do on a marginal level, which are covered by Tomasik (2015c).

I argue that, despite astronomical waste, we should spend more time determining the expected value (magnitude and sign) of the future, and whether we can improve it, before accelerating progress. We should seek to determine the best values to spread and how to spread them before doing so, and only once we have sufficiently spread good values should we accelerate progress. Prior to all of this, however, is the need to determine whether there is a true moral theory, and if so, what it is, and if there isn't, what we should do. The answers to these questions dictate all else.

High impact research questions

This section proposes a number of high value research questions have not yet been explored, or would benefit from further research.

1) Terraforming related questions (covered above in 'Terraforming Mars'). How much wild-animal life could be sustained on Mars were it to become partially habitable?

2) AI safety related questions (see Bostrom 2014 for a complete summary of unanswered questions in this field).

3) Whether digital systems (and fundamental physics, as proposed by Tomasik (2016f)) can be sentient and/or experience suffering.

It is often assumed that digital systems will be sentient, or that it is highly likely (e.g. Bostrom 2014). However, while I am an outsider to the field of cognition and AI, I don't believe that this is obviously so. It is argued that digital computation is not enough for sentience, for example in Searle's Chinese Room thought experiment (Searle 1980), and by O'Brien & Opie (2001). Therefore, I suggest that determining whether digital systems can be sentient is a critical task for understanding the implications of an intelligence explosion. If it were the case that digital systems could not be sentient and they were propagated throughout the universe (instead of biological minds), this would result in a dark universe indeed, full of mindless robots and software acting like sentient minds, but experiencing nothing.

4) Tomasik (2016g) has suggested a survey on the effects of advocating veganism (particularly with an abolitionist or rights based messaging) on people's willingness to intervene in nature.

If convincing people to become vegan makes them less likely to want to alter the environment to reduce suffering (or spread wellbeing throughout the universe), this could be very bad, and may outweigh the short term benefits of more vegans. The impact of becoming vegan on one's disposition towards intervening in the environment likely depends on the line of messaging used to convert them.

5) What is the probability that the far future will be good/bad, and how good/bad might it be? What is the expected value of the future? Is the worst possible outcome equal in magnitude but opposite in sign to the best possible outcome?

Many assume that the future, barring some existential catastrophe, will be, on balance, good. This is not obvious, especially once we consider non-human minds. Further work should be undertaken to determine the amount of suffering and wellbeing there will be on expectation.

6) Can we compare and aggregate suffering and wellbeing?

Some have argued that it does not make sense to be able to aggregate the suffering and wellbeing of a group of sentient beings without suffering and wellbeing being objectively measurable to some high degree (e.g. Knutsson 2016), and that it is impossible to determine how many 'units' of wellbeing cancels out some amount of suffering.

To answer this question is beyond the scope of this essay, but I seek here to highlight that it is not obvious that we cannot aggregate suffering and wellbeing. The same neural structures that produce negative valence also produce positive valence, which are associated with suffering and wellbeing respectively for conscious minds. Further, there are many instances of psychologists measuring pain and wellbeing in objective ways that do not require value judgements. For further reading on this discussion, see the work of Daniel Kahneman and Dan Ariely (e.g. Kahneman et al 1997).

This is a critical question to answer for determining whether it makes sense to ask the question of whether the expected value of the future is positive or negative from an objective basis, or whether everything is simply based on value judgements.

6) Is there a true moral theory? If so, what is it? What are the best values?

7) Modelling the effects on the far future of values spreading and other present day interventions.

8) How do we best spread good values?

9) How strong is the pro-natal bias at leading conscious minds to believing that their life is worth living? Can we correct for this when determining whether an individual is, on balance, experiencing more wellbeing than suffering?

Caution should be taken in pursuing these research avenues, as many are directly applicable to advancing the field they seek to determine the implications of. For example, it is conceivable that researching the likelihood or impacts of terraforming will advance terraforming science generally and speed it up, regardless of whether it is determined to be, on expectation, a good thing to do so. This risk must be carefully weighed, and this essay does not have a good answer for how to do this.

Conclusion

This essay sought to provide an overview of the literature relevant to wild-animal suffering, terraforming and the far future. Suffering of wild animals and invertebrates in the wild is likely a large source of pain, and spreading wild animals to other planets is expected to be astronomically bad. Even the risk of this dictates extreme caution. Some of the ethical considerations important for discussing wild-animal suffering were also covered, and some new insights were offered. In particular, some recommended actions and a research agenda were proposed. Some key conclusions of the essay are outlined below.

- Discussion of the best underlying philosophy is critical, as several different ethical codes (including negative and classical hedonistic utilitarianism) each arrive at different answers to the question of what to do about the far future.
- Without AGI, terraforming of Mars and the spreading of wildlife to other planets may be possible in 150 years. It is highly likely, but not a foregone conclusion, that AGI will reach an intelligence explosion by that point.
- If Mars is terraformed, it is plausible that it can eventually become home to almost as much wild-animal suffering as there currently exists on Earth's land.
- Values spreading is one of the most high impact ways to positively impact the far future, although we first need to be confident we are spreading the best values.
- There is a limited amount of time for solving the value spreading problem for spreading wild-animal suffering, e.g. encouraging concern for wild animals, utilitarianism (or otherwise finding the true or best moral theory given normative uncertainty), and spreading concern for spreading wellbeing. These problems are also critical for determining what values to load to an AGI.
- I have proposed some reasons for why person-affecting views and negative utilitarianism may be flawed and argue in favour of classical hedonistic utilitarianism though I am not 100% certain about this (nor will I ever be, due to normative uncertainty), and this is meant to create dialogue as well as to criticise.
- We will never be 100% certain that we have identified the best values, and therefore we should consider how certain we want to be before we switch to primarily focusing on spreading values. Once the majority of society has values that we believe are best with some degree of certainty, we can then focus further on ensuring that the values we have chosen are the best ones. A thorough investigation of this is well beyond the scope of this essay, but is strongly called for.

Some of the conclusions of this essay are tentative, and would benefit from significantly more consideration and research. This essay was meant to suggest some solutions and insights to important questions and encourage discussion.

I argue for caution towards terraforming Mars or otherwise colonising space due to the risk of spreading wild-animal suffering (or suffering in general in the long term), and instead I recommend undertaking high impact research to determine the expected value of the future. I also strongly urge discussion to determine the best ethical theory, and then to determine the best values to spread for that theory, followed by researching the best ways to spread them, and finally enacting on their spreading.

Surely (assuming I am right about the normative issues), the best outcome is spreading the maximum possible wellbeing throughout the universe, with the worst outcome being spreading the maximum possible suffering. These are the realisation of Sam Harris' best and worst possible worlds. We are in position now to set up the future such that the best possible world is a reality, and it is imperative that we do so. Nothing else, save perhaps ensuring that there is a future for sentience, is more important.

References

Beckstead, N. 2013, On the overwhelming importance of shaping the far future, PhD dissertation, Rutgers University-Graduate School-New Brunswick.

Benetar, D. 2008, Better never to have been: The harm of coming into existence, Oxford University Press.

Bostrom, N. 2003, Astronomical waste: The opportunity cost of delayed technological development, Utilitas, V. 15, 308-314.

Bostrom, N. 2013, Existential risk prevention as global priority, Global Policy, V. 4, 15-31.

Bostrom, N. 2014, Superintelligence: Paths, dangers, strategies, OUP Oxford.

Brown, D., Cantillo, L., Neal-Jones, N., Steigerwald, B. & Scott, J. 2015, NASA mission reveals speed of solar wind stripping Martian atmosphere, 6 November 2015, <http://www.nasa.gov/press-release/nasa-mission-reveals-speed-of-solar-wind-stripping-martian-atmosphere> (28 August 2016).

Dickens, M. 2015, On values spreading, Philosophical Multicore, 10 September 2015, <http://mdickens.me/2015/09/10/on_values_spreading/> (20 August 2016).

Dickens, M. 2016a, Quantifying the far future effects of interventions, Effective Altruism Forum, 18 May 2016, <http://effective-altruism.com/ea/xq/quantifying_the_far_future_effects_of/#animal-advocacy> (24 August 2016).

Dickens, M. 2016b, The myth that reducing wild animal suffering is intractable, Philosophical Multicore, 22 April 2016, <http://mdickens.me/2016/04/22/the_myth_that_reducing_wild_animal_suffering_is_intractable/> (20 August 2016).

Dyson, F.J. 1960, Search for artificial stellar sources of infra-red radiation, Science, V. 131, 1667-1668.

Fogg, M.J. 2011, Terraforming mars: A review of concepts, Advances in Space Research, V. 22, 2217-2225.

Graham, J.M. 2004, The biological terraforming of Mars: Planetary ecosynthesis as ecological succession on a global scale, Astrobiology, V. 4, 168-195.

Hanson, R. 1998, Burning the cosmic commons: Evolutionary strategies for interstellar colonization, <http://mason.gmu.edu/~rhanson/filluniv.pdf>.

Hanson, R. 2009, Limits to growth, Overcoming Bias, 21 September 2009, <http://www.overcomingbias.com/2009/09/limits-to-growth.html> (24 August 2016).

Harris, S. 2011, The moral landscape: How science can determine human values, Simon and Schuster.

Kahneman, D., Wakker, P.P. & Sarin, R. 1997, Back to Bentham? Explorations of experienced utility, *The Quarterly Journal of Economics*, V. 112, 375-405.

Knutsson, S. 2016, How could an empty world be better than a populated one?, Foundational Research Institute, 24 August 2016, <https://foundational-research.org/how-could-an-empty-world-be-better-than-a-populated/> (24 August 2016).

MacAskill, W. 2014, Normative uncertainty, PhD dissertation, University of Oxford.

Madhusoodanan, J. 2014, Microbial stowaways to Mars identified, Nature, 19 May 2014, <http://www.nature.com/news/microbial-stowaways-to-mars-identified-1.15249> (22 August 2016).

Matheny, J.G. 2007, Reducing the risk of human extinction, *Risk Analysis*, V. 27, 1335-1344.

Marinova, M.M., McKay, C.P. & Hashimoto, H. 2005, Radiative-convective model of warming Mars with artificial greenhouse gases, Journal of Geophysical Research: Planets, V. 110.E3.

McAdam, D. 1986, Recruitment to high-risk activism: The case of Freedom Summer, American Journal of Sociology, V. 92, 64-90.

McKay, C.P. 1990, Does Mars have rights? An approach to the environmental ethics of planetary engineering, in: D. MacNiven (Ed.), Moral Expertise, Routledge.

McKay, C.P., Toon, O.B. & Kasting, J.F. 1991, Making Mars habitable, Nature, V. 352, 489-496.

McKay C.P. 2007, Planetary ecosynthesis on Mars: Restoration ecology and environmental ethics, NASA Ames Research Center.

Müller, V.C. & Bostrom, N. 2016, Future progress in artificial intelligence: A survey of expert opinion. In Fundamental issues of artificial intelligence, Springer International Publishing, p. 553-570.

Murphy, T. Galactic-scale energy, Do the Math, 12 July 2011, <http://physics.ucsd.edu/do-the-math/2011/07/galactic-scale-energy/> (24 August 2016).

O'Brien, G. & Opie, J. 2001, Connectionist vehicles, structural resemblance, and the phenomenal mind, Communication and Cognition, V. 34, 13-38.

Pearce, D. 2004, The Hedonistic Imperative, David Pearce.

Rees, M. 2003, Our final hour: a scientist's warning, Basic Books.

Sandberg, A. 2009, The hot limits to growth, Andart, 23 September 2009, <http://www.aleph.se/andart/archives/2009/09/the_hot_limits_to_growth.html> (24 August 2016).

Schwartz, J.S.J. 2013, On the moral permissibility of terraforming, Ethics and the Environment, V. 18, 1-31.

Searle, J.R. 1980, Minds, brains, and programs, Behavioral and Brain Sciences, V. 3, 417-424.

Shishko, R., Fradet, R., Saydam, S., Tapia-Cortez, C., Dempster, A.G. & Coulton, J. 2015, An integrated economics model for ISRU in support of a Mars colony - initial status report, AIAA Space 2015 Conference and Exposition.

Sparrow, R. 2015, Terraforming, Vandalism and Virtue Ethics, in: J. Galliott (Ed.), Commercial space exploration: Ethics, policy and governance, Routledge.

Tomasik, B. 2015a, The importance of wild-animal suffering, Relations. Beyond Anthropocentrism, V. 3, 133-152.

Tomasik, B. 2015b, Will space colonization multiply wild-animal suffering? Reducing Suffering, 12 March 2015, <http://reducing-suffering.org/will-space-colonization-multiply-wild-animal-suffering/> (20 August 2016).

Tomasik, B. 2015c, Differential intellectual progress as a positive-sum project, Foundational Research Institute, 21 December 2015, <https://foundational-research.org/differential-intellectual-progress-as-a-positive-sum-project/> (25 August 2016).

Tomasik, B. 2016a, Risks of astronomical future suffering, Foundational Research Institute, 5 July 2016, <https://foundational-research.org/risks-of-astronomical-future-suffering/> (20 August 2016).

Tomasik, B. 2016b, How many wild animals are there?, Reducing Suffering, 7 May 2016, <http://reducing-suffering.org/how-many-wild-animals-are-there/> (25 August 2016).

Tomasik, B. 2016c, Do bugs feel pain?, Reducing Suffering, 27 April 2016, <http://reducing-suffering.org/do-bugs-feel-pain/> (28 August 2016).

Tomasik, B. 2016d, The importance of wild-animal suffering, Foundational Research Institute, 22 April 2016, <https://foundational-research.org/the-importance-of-wild-animal-suffering/> (20 August 2016).

Tomasik, B. 2016e, Convert grass lawns to gravel to reduce insect suffering, Reducing Suffering, 25 August 2016, <http://reducing-suffering.org/convert-grass-lawns-to-gravel-to-reduce-insect-suffering/> (27 August 2016).

Tomasik, B. 2016f, Is there suffering in fundamental physics?, Reducing Suffering, 10 June 2016, <http://reducing-suffering.org/is-there-suffering-in-fundamental-physics/> (15 August 2016).

Tomasik, B. 2016g, Does the animal-rights movement encourage wilderness preservation?, Reducing Suffering, 5 August 2016, <http://reducing-suffering.org/does-the-animal-rights-movement-encourage-wilderness-preservation/> (19 August 2016).

United Nations 2002, United Nations treaties and principles on outer space, <http://www.unoosa.org/pdf/publications/STSPACE11E.pdf>.

Von Neumann, J. & Burks, A.W. 1966, Theory of self-reproducing automata, IEEE Transactions on Neural Networks, V. 5, 3-14.

Xie, J., Sreenivasan, S., Korniss, G., Zhang, W., Lim, C. & Szymanski, B.K. 2011, Social consensus through the influence of committed minorities, Physical Review E, V. 84.1: 011130.

Yamashita, M., Hashimoto, H. & Wada, H. 2009, On-site resources availability for space agriculture on Mars, in: Mars: prospective energy and material resources, Springer Science & Business Media.

Zubrin, R., Baker, D. & Gwynne, O. 1991, Mars Direct: A simple, robust and cost-effective architecture for the space exploration initiative, AIAA 91-0326, 29th Aerospace Science Conference, Reno, NY.

Zubrin, R. 2011, The case for Mars: The plan to settle the red planet and why we must, Simon and Schuster.