ARTICLE IN PRESS

Space Policy xxx (xxxx) xxx

Contents lists available at ScienceDirect

Space Policy

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

Discounting, Buck-Passing, and Existential Risk Mitigation: The Case of Space Colonization

Joseph Gottlieb

Texas Tech University, United States

ARTICLE INFO

Article history:
Received 19 June 2021
Received in revised form
22 November 2021
Accepted 13 February 2022
Available online xxx

Keywords:
Space colonization
Existential risks
Suffering risks
Discounting
Buck-passing
Long-term future

ABSTRACT

Large-scale, self-sufficient space colonization is a plausible means of efficiently reducing existential risks and ensuring our long-term survival. But humanity is by and large myopic, and as an intergenerational global public good, existential risk reduction is systematically undervalued, hampered by intergenerational discounting. This paper explores how these issues apply to space colonization, arguing that the motivational and psychological barriers to space colonization are a special—and especially strong—case of a more general problem. The upshot is not that large-scale, self-sufficient space colonization will never occur, but that, absent institutional change, the conditions under which it is most likely to occur are precisely those conditions where the threat of suffering risks might be most high.

1. Existential risk mitigators

Let an *existential risk* ('x-risk') be any risk that threatens the destruction of our long-term potential, either through literal extinction or an unrecoverable collapse (Bostrom, 2013); [6,36]. Examples of x-risks include unaligned artificial intelligence, nuclear war, runaway climate change, asteroids, molecular nanotechnology, and engineered pandemics. Apart from asteroid strikes, these threats are relatively new, and entirely anthropogenic; as such, our longevity as a species should give us little hope for our survival over the long term [6, p. 16].

Let an *x-risk mitigator* be anything that reduces an x-risk in a meaningful way. Even with 'meaningfully' left purposely undefined, there are plenty of clear cases: nuclear non-proliferation treaties are an x-risk mitigator for nuclear war, research into the value-alignment problem is an x-risk mitigator for artificial intelligence, and reduction of fossil fuel consumption is an x-risk mitigator for climate change. The focus of this paper is on space colonization as an x-risk mitigator—specifically, impediments to using space colonization to mitigate x-risks.

Impediments to space colonization are well-known, but the tack taken up here is relatively unique. One commonly considered set of impediments are *physical impediments*: impediments that are a function of human-independent facts about the world and its underlying nature. A physical impediment to space colonization might be the lack of extraterrestrial water sources. Whether or not this is an actual impediment—as it happens, it does not seem to be—the presence of extraterrestrial water sources is not contingent on any human-dependent facts. There are also *technological impediments*: impediments that are a function of our current technological know-how and prowess. Even if there were extraterrestrial water sources, our not being able to access those sources would be a technological impediment.[1]

I focus on *psychological impediments*. But I am not interested in the usual suspects, e.g., depression due to being far from home. This

[1] While not the focus of this essay, physical and technological impediments are serious. In the case of extra-terrestrial water extraction (a technological impediment), one recent proposal stems from the Mars Rodwell Experiment Final Report, which documented several tests from The Aerospace Corporation, investigated the use of a 'Rodriquez Well', which was first tested at the Amundsen—Scott South Pole Station in the 1970s. The basic process involves drilling through the layers of dust and rock to reach ice deposits and using a submersible pump to create a cavity filled with water, which (after being heated) is returned to the surface. The report from NASA, which includes a discussion of the challenges of carrying over the Rodriquez Well technology to a Martian environment, can be found in https://ntrs.nasa.gov/api/citations/20205011353/downloads/Rodwell%20Experiment%20Final%20Report%20TP-20205011353.pdf.

E-mail address: Joseph.gottlieb@gmail.com.

https://doi.org/10.1016/j.spacepol.2022.101486
0265-9646/

Please cite this article as: J. Gottlieb, Discounting, Buck-Passing, and Existential Risk Mitigation: The Case of Space Colonization, Space Policy, https://doi.org/10.1016/j.spacepol.2022.101486

and related issues have been examined extensively (e.g., Refs. [1,30]), but they do not constitute impediments to space colonization *as such*; they constitute impediments to the colonization effort going well for the colonists. Instead, I will consider the following:

(i) Our tendency toward steep positive intergenerational discounting;

And, to a lesser extent:

(ii) Our tendency toward intergenerational buck-passing.

These psychological impediments have wide-spread application to x-risk mitigation in general and short-circuit our ability to preserve the future [36]. Their role as impediments to space colonization is thus a special case of a more general problem, but is, in my estimation, a relatively underexplored special case. And if I am right, it is in many ways a worse case, as the barriers to space colonization (i) and (ii) present are particularly strong and cannot easily be assuaged by standard countermeasures.

Let me be more precise about my aims, the scope of my thesis, and the assumptions being made. First, my aim is to explore the ways in which (i) and (ii) adversely impact the prospects of building permanent, self-sustaining, large space colonies (i.e., space colonies suitable for x-risk mitigation) by space-faring states *where ordinary citizens have some measure of legislative control and legislature members operate on short election cycles*. Here, 'ordinary' is not employed as a pejorative; 'ordinary', in the relevant sense, means most people—people who are mostly worried about dealing with the trials of today or tomorrow, not what may happen decades from now. What counts as 'some measure of legislative control' can merely be indirect control. So, in both direct democracies (as in Switzerland) and representative democracies (as in the United States), citizens have some measure of legislative control; it is just that in the latter it is indirect via the election of representatives who—in the ideal case—reflect the preferences of their constituents. As it happens, this characterization obtains, for better or worse, in most, though not all, space-faring nations.

Second, while the operative forms of discounting and buck-passing and the devaluing of the future they reflect are deeply immoral—more on this below—my stance, at least in the first instance, is to take these as descriptive facts described in humanity in general. That is, they accurately describe how most people *are*, not necessarily how we would *like* them to be.

An example. Parfit [38] has us considered the following three possibilities concerning humanity's future:

(1) There is no catastrophe.
(2) There is a catastrophe that immediately kills 99% of the world's population.
(3) There is a catastrophe that immediately kills 100% of the world's population.

When comparing each outcome, we tend to think that (2) is worse than (1) and that (3) is worse than (2). In a recent study conducted by Schubert *et al.* [44] covering lay participants across the United States and United Kingdom, this is indeed what was found, with 83% of participants ranking (1) as the best outcome and (3) as the worst outcome. This was the 'expected ranking'. But Schubert *et al.*'s primary interest was elsewhere. Of those who gave the expected ranking, they asked which difference they felt was greater: the difference between (1) and (2) or the difference between (2) and (3)? In other words, they asked participants if extinction is *uniquely bad*. What Schubert et al. found was that, for the most part, people think the answer is 'no': only about 23.5% judged the difference between (2) and (3) to be greater than the difference between (1) and (2). This betrays a misguided attitude about our extinction. What matters is not just the present lives that may be cut short; what also matters—and indeed what may matter more—is that, in expectation, almost everything of value lies in the future [36]. So, x-risk mitigation is a future-oriented enterprise, and thus, space colonization for x-risk mitigation is itself a future-oriented enterprise—especially so, as we will see, considering what a space colony must be like if it is meant to reduce x-risks. It is for this reason that discounting and buck-passing are emphasized, as they reflect our widespread myopia.[2]

Here is the plan. I begin by briefly discussing why space colonization should be pursued as an x-risk mitigator in the first place (Section 2). Having done so, I offer an extended discussion of the psychological impediments themselves (Section 3) and why, when it comes to space colonization, they are especially hard to surmount (Section 4). Finally, I conclude (Section 5) by showing that the exceptions—i.e., cases where (say) discounting is *not* operative—while perhaps making space colonization more likely, may nonetheless result in space colonies that are suboptimal *for the colonists*. We do not simply want space colonization to happen. We also want space colonization to go well. Several policy recommendations are then set out in light of this result.

2. Safeguarding the future via space colonization

From the perspective of x-risk mitigation, there is an intuitive case for space colonization that trades on space colonization creating *catastrophe-independent* pockets of humanity. Let a 'pocket' of humanity be catastrophe independent from another pocket if a catastrophic event in one pocket does not necessarily cause a catastrophic event in another. Then, we can argue like this:

SC1: For a catastrophic event E to cause either civilizational collapse or literal extinction, every pocket of humanity must be impacted by E.
SC2: By making the pockets of humanity catastrophe independent, we vastly lower the likelihood that every pocket of humanity is impacted by E.
SC3: Space colonization would make the pockets of humanity catastrophe independent.
∴ Space colonization will vastly lower the likelihood that every pocket of humanity is impacted by E.

This argument is intuitive as it trades on the adage that we should not put all our eggs in one basket (when it comes to our long-term survival). When SpaceX's Elon Musk urges that we "back up the hard drive" by colonizing Mars, he is colorfully pointing to something like this idea.[3] The core point is this: if the conclusion of this argument is true, then space colonization is a good x-risk mitigator.

Of course, this argument faces several challenges, with SC3 being the obvious target. One worry is risk correlation. To what extent are the various x-risks statistically correlated across

[2] Note that these facts are especially concerning given that is plausible that—despite some low-risk exceptions, like stellar explosions—humanity can do something to mitigate x-risks. In this respect, we are unlike the other many terrestrial species that have gone extinct. Indeed, we are likely the only species that even has the concept of x-risks. Larry Niven once said that "dinosaurs became extinct because they didn't have a space program. And if we become extinct because we don't have a space program, it'll serve us right!". Niven was not wrong.

[3] See https://www.gq.com/story/elon-musk-mars-spacex-tesla-interview for this line.

settlements? [36, p. 194]. If one risk impacts Earth, how likely will it be that it impacts (e.g.) a colony on Mars? It is challenging to answer this question in the abstract without knowing what the off-Earth space colonies might be like—a point we will broach later. Perhaps some x-risks, like those associated with misaligned artificial super intelligence, will be strongly correlated across settlements. Even so, all that shows is that space colonization is not sufficient to secure our long-term future. That is consistent with it removing the uncorrelated portion of our total x-risk and thus significantly increasing our chances of long-term survival [36, p. 394].[4]

Another worry trades on the idea that x-risk mitigation is a *knapsack problem*. The overall goal of minimizing x-risk is likely best conceptualized as a matter of picking the right range of projects given a certain budget, and space colonization (we might say) is just too expensive given its benefits. So, while SC3 might be true, this does not mean we should pursue space colonization for x-risk mitigation [4]. Perhaps there are other means of mitigating x-risks that cost less and will lower x-risks at least as much as space colonization. Baum *et al.* [2, p. 53] take something like this line, suggesting that the main point in favor of terrestrial refuges over extraterrestrial refuges is their relative costs, remarking that “if the only goal is to minimize the risk of global catastrophe, then extraterrestrial refuges are unlikely to be advantageous”.

Terrestrial-based refuges—even if they, as is likely, are subterranean or ocean-based—do not protect against all existential threats (e.g., massive asteroid impacts), and it is at best unclear if there are any threats for which terrestrial-based refuges provide better protection than extraterrestrial refuges. But put that aside: while space colonization will indeed be expensive, whether its costs stand up to its benefits as an x-risk mitigator is an open question. For one thing, much of this depends on technological progress, a point that Baum *et al.* grant [2] and here, there have been some inroads. As of 2018, SpaceX's Falcon 9 platform has reduced the cost to LEO by a factor of twenty [29], to $2720/kg. We have no reason to doubt that this figure will only improve, even if it falls far short of the $10 USD/kg aspirational mark for SpaceX's Starship. The cost of space colonization will naturally vastly exceed LEO launch costs, but since getting to LEO is a requirement for *any* colonization effort, this is a reason for hope.

To sum up, I do not take the considerations adduced here to present a conclusive case for space colonization as a means of x-risk mitigation. But I do hope that they are strong enough to warrant considering this paper's central question: why efforts toward space colonization face unique challenges from discounting and buck-passing. So going forward, I will assume that space colonization should be pursued for existential security. On the assumption that the preservation of our future is of deep moral importance and indeed obligatory, then it follows that one of the most important things we can do (space colonization) to fulfill one of our most important moral obligations (preserving our future), faces steep obstacles.[5]

3. The Impediments

3.1. Intergenerational discounting

Intrapersonal hedonic discounting—discounting positively or negatively valanced events within our own lives—is often based on a pure time preference, where the time preference is ‘pure’ insofar as we prefer one benefit over another *simply because* of its temporal proximity. Here are two examples:

> Good news—you have won a radio contest! There are two possible prizes, and the announcer is about to make the big reveal. One prize is a pleasant skiing vacation in Vermont to take place this weekend. The other prize is a lavish, week-long ski trip in the Swiss Alps five years from now. Suppose that you do not have any anticipated scheduling conflicts and that your prize—whichever it is—is assured and nontransferable. Which do you prefer? Many of us can handle some delayed gratification, but some may think five years is just too long to wait for a good thing [19, p. 947].

> Before you lay the cookies; they are very tasty. You can either have one of these cookies tomorrow or two of them next week. Which do you prefer? In isolation, having two cookies is better, but is it better enough to call for waiting an entire week? You are tempted to choose one cookie because of the temporal nearness of its benefit. The two cookies involve far more waiting [19].

In essence, this form of discounting amounts to impatience. It is thus different from discounting future monetary benefits (an instrumental good) due to diminishing marginal utility.

Now the benefits of large-scale space colonization can be hedonic. For those who lay the necessary groundwork for a colony—those who, say, take the first steps on Mars and lay down the infrastructure for a future colony—it may well be thrilling. For the rest of us though—people alive now who might prefer a world in which space colonization is prioritized yet not actually partake in (e.g.) setting up the initial Martian infrastructure—the benefits are non-hedonic. And that's what's operative here. We will not live in a self-sustaining Martian colony. Indeed, it is unlikely that our children will live to see its full fruition. So, when we decide to pursue space colonization now, we are undertaking a project whose realization will benefit future generations. The discounting is not intrapersonal. Rather, it is *intergenerational*.

We are not considering space colonization as a new form of tourism for the wealthy; we are considering space colonization as a means of protecting our future—as a means of reducing x-risks. So, the future benefit in question is that of there being any future and future well-being at all. In this way, when considering the discounting of well-being itself, we are asking how much (if at all) future lives matter and whether the value of well-being depends on when (relative to the present) it occurs. And since in expectation almost every value lies in the future—absent an extinction event, most generations are future generations—what is being weighed in intergenerational trade-offs concerning well-being is benefits limited to the present or (very) near-term generations versus those that are applicable for all generations to come. To have a positive discount rate on future well-being, x amount of *our* well-being *now* is worth more than that same amount of *future* well-being for *future* people. At a (fairly common) 5% positive

[4] The removal of uncorrelated risk via space colonization also plausibly trumps any damage to celestial objects qua objects of scientific inquiry that would inevitably be caused by space colonization. And since scientific interests degrade over time, principles of scientific conservation cannot justifiably delay colonization forever anyway. For discussion, see Ref. [21].

[5] Precisely why it is important is a matter of controversy. For different takes, see Refs. [14,16,36]. A key point of division is whether extinction matters only for the people who perish in the extinction event itself or whether it matters also because of the loss of an immense amount of expected future value. I am implicitly assuming the latter view here, but it should be noted that even if the former view is correct, extinction would still be an immense tragedy, given that our current population verges on 7.8 billion, with an average life expectance of over 72 years. For discussion, see Ref. Beckstead (2013) [5].

"MATCHING"

Each year some people in the United States may die as a result of exposure to certain kinds of pollutants. Unless there are programs to control this pollution, some will die this year from pollution, and some will die in the future. Consider **Program A** and **Program B** below, and then fill in the blank such that you would be indifferent between them—so that you would judge them to be equally good.

Program A will save 100 lives this year, but will save 0 lives 100 years from now.
Program B will save 0 lives this year, but will save ___ lives 100 years from now.

(Please look back at the number you wrote in the blank to make sure that you consider Policy A and Policy B to be equally good. If you change your mind, feel free to change your answer.)

Fig. 1. Matching task. Reproduced from the study by Frederick [15, p. 42].

discount rate for future well-being, the value of someone's well-being in 100 years is only 0.76% of the well-being of a present person, and a death in 75 years will only be 2.58% as significant as a death today.[6] From this perspective, it is easy to see why one might balk at large-scale self-sufficient space colonization: why would I pursue space colonization over a near-term project when the beneficiaries of the former are worth so few?

Now there are at least two questions we might ask at this point. First, is this kind of discounting ethically defensible? And second, do we, at least in general, exhibit this sort of discounting behavior? On the first question, discounting future well-being based is almost certainly indefensible, at least if based on purely temporal proximity. As Cowen and Parfit [10] point out, it leads to absurd consequences, like your death by cancer at 25 being justified because Cleopatra wanted an extra helping of dessert. But in the first instance, this does not matter here. What matters is our second question. For it is our actual behavior, not what our behavior should be, that will determine how we will act (or not act) to mitigate an x-risk. Even if *prescriptivists* are correct that intergenerational discount rates should be based on ethical considerations—and we will consider how ideas favorable to prescriptivism might be implemented at an institutional and policy level later on—what drives the problem is *descriptive* facts about our revealed preferences. And it turns out that we *do* discount future well-being.

Using a telephone survey, Cropper, Aydede, and Portney (1992) [11] asked respondents to choose between two life-saving programs: one of which would save a specified number of lives now and the other of which would save a far greater number in the future. They found that respondents would trade 45 lives in 100 years to save 1 life today. Johannesson and Johansson [26] found even more drastic substitution rates: up to 243 to 1 for a similar time horizon.

Of course, these data points are not without controversy. Frederick [15] notes that the degree of intergenerational preferences is sensitive to elicitation procedures. Using a 'matching' task (Fig. 1), Frederick found a median matching response of 324, implying a median substitution rate of 3.2 to 1. This figure is vastly lower than that found by Cropper, Aydede, and Portney (1992) [11] and Johannesson and Johansson [26], both of which used 'choice' paradigms (Fig. 2).

Frederick [15] speculates that the matching task has lower substitution rates than choice tasks as it frames the comparison being made in less loaded terms:

> Asking respondents to compare a program that saves a smaller number of lives now with a program that saves a greater number of lives in a future year compels the inference that there must be some reason to discount the future lives—otherwise, why would the experimenter be asking the question. The matching task has no such demand effect because the intertemporal options are not fully specified; the respondents must generate the indifference value themselves.

In response, I will make three points. First, we can allow that not everyone exhibits intergenerational discounting. Indeed, as we will see, the current story predicts that if large-scale space colonization is to be possible at all, there must be people who do not exhibit intergenerational discounting.[7] So Frederick's results are inconsistent with the present thesis only insofar as they establish that most of the populace exhibits intergenerational neutrality or (less likely) a negative discount rate on future well-being.

Second, Frederick's results notwithstanding, the literature on discounting in the context of x-risks illustrates how intergenerational discounting hampers collective efforts. For example, Jacquet *et al.* [25] showed that in collective-risk games to avoid "dangerous climate change", intergenerational discounting was the central factor undermining cooperation. Since climate change is an x-risk and we are considering space colonization *qua x-risk mitigator*, it is not a stretch to apply Jacquet *et al.*'s [25] results here.

Third and most importantly, Frederick [15, p. 40] is explicit that his descriptive research concerns temporal intergenerational discounting. Yet it is not innocuous to assume that intergenerational trade-offs operate on a time-preference model to begin with, and no such assumption has been made here. A time-preference model of intergenerational trade-offs would predict that the non-hedonic

[6] To see this, suppose we were able to quantify well-being and that a philanthropist P in the present is trying to decide whether to immediately 'distribute' some number of units of well-being (perhaps by financing some charitable intervention) or defer this benefit to some far future people. Suppose further that P employs a 5% discount rate per year. At a 5% rate, P is saying that well-being had by people in one year's time is worth 5% less than the very same amount of well-being had by people now, such that for P to be neutral between a present benefit of 100 units of well-being and (e.g.) a well-being benefit received in one year, that future benefit must be roughly 105 units of well-being. We can then compute the *present value* ('*PV*') of this future benefit via the following formula: $PV = FV/(1 + r)$, where '*FV*' picks out the future value, '*r*' picks out the discount rate, and '*t*' picks out the time horizon. This means that at a 5% discount, 100 units of well-being (*FV*) in the 100 years (*t*) are only worth 0.76 of a unit of well-being received now. Discount rates are standardly computed using the Ramsey [41] formula: $r = \delta + ng$, where '*g*' denotes the growth rate of consumption, '*n*' denotes the elasticity of the marginal utility of consumption, and 'δ' denotes our rate of pure time preference.

[7] In this way, I agree with Haqq-Misra [22] (p. 152) that "deep altruism"—understood as "selfless concern for the well-being of others in the distant future"—may be required to sustain space settlement.

"CHOICE"

Each year some people in the United States may die as a result of exposure to certain kinds of pollutants. Unless there are programs to control this pollution, 100 people will die this year from pollution, and 7000 people will die 100 years from now. The government has to choose between two new programs to control this pollution. The two programs cost the same, but there is only enough money for one.

Program A will save 100 lives now.
Program B will save 7000 lives 100 years from now.

Which program would you choose? (circle one) **A** **B**

Fig. 2. Choice task. Reproduced from the study by Frederick [15, p. 41].

valuing involved in space colonization exhibits near bias akin to the vacation and cookie cases. It would predict that for any future benefits B1 and B1, where B2 is at least as valuable as B1, and adjusting for the relevant subjective probabilities of their occurrence, we would prefer B1 precisely because it would occur nearer to us in time. Assuming we do discount in this way, given the time scales involved in space colonization, we should expect that there are numerous benefits we would prefer over space colonization simply because of their temporal proximity.

But it is not at all obvious that the time-preference model is apt for non-hedonic valuing, including the non-hedonic valuing inherent in space colonization. An immediate reason to be suspicious of its aptness is that, technically, time preferences only correspond to experienced time. Greene [19] illustrates this nicely with the following case:

> Tomorrow morning you will be put into a consciousness-free stasis for a year. You have a choice between the \$50 meal this evening and the \$100 meal after you exit the stasis. You know that it would feel as if the \$100 meal occurs tomorrow.

Setting aside worries about the stasis lowering the probability that you will get to eat the \$100 meal, cases like this reveal that time-preference models of hedonic valuing assume that subjective time reflects objective time. Fundamentally, the former is doing the work. But if that is so, temporal discounting will not apply to intergenerational trade-offs, since intergenerational tradeoffs do not involve subjective time ([19]).

For this reason and others, Greene [20] suggests that we ought to model intergenerational trade-offs in terms of *perceived social distance*. This is a form of social discounting where, given two exclusive future benefits B1 and B1, we would prefer B1 *precisely because* B1 concerns people who are of greater perceived social closeness, like kinship. And it is obvious that, just as with the time-preference model of intergeneration trade-offs, a social distance model will adversely impact the prospects of space colonization as the would-be colonizers will not just be temporally distant, they will be socially distant as well.[8] There is nothing in Frederick's results that suggest that we do not discount by perceived social distance.

Before we press on, there is a further point worth noting: the structures and practices of governments are overwhelmingly myopic (temporally and socially), and so will work in ways that favor the short term, and thus mirror the impacts of intergenerational discounting. Part of this mirroring will of course just be reflective of the underlying individual political actors themselves possessing positive discount rates.[9] But there are also institutional drivers like election incentives that deliver the same results. Politicians want to be re-elected. Yet because they operate on—and are rewarded in terms of—short time horizons, they are disincentivized to focus on projects (like space colonization) that are inherently intergenerational.

3.2. *Buck-passing*

We spend comparatively little to mitigate existential risks. As Ord [36, pp. 57–58] (fn. 54, 55) laments, the annual budget for the international body that is responsible for bioweapon prohibitions is \$1.4 million (less than the average McDonald's budget), whereas total global spending on reducing existential risk from artificial intelligence is no more than \$50 million and perhaps as little as \$10 million.

This is not surprising, even if it is disappointing. Mitigating existential risk is a *global intergenerational public good* [36, pp. 58–59]. This has two consequences of note: first, the global side. Any party that works to mitigate existential risk will bear the full cost but reap only a small percentage of the benefit. Currently, the United States constitutes 4.5% of the world's population. As such, if the United States government–implemented policies gear toward existential risk mitigation, the benefits would be largely enjoyed by other countries [36]. But of course, this point is wholly general, since outside of China and India, no nation currently constitutes more than 4.5% of the world's population. Thus, we should expect most if not all nations to undervalue such mitigation efforts and prefer to free ride—that is, opt to enjoy the spillover effects while another country bears the financial costs (cf. [23]).

The second problem is more to the point. For even if we could achieve global burden-sharing for existential risk mitigation, there would still be a temptation to undervalue, since the beneficiaries include all the people that might ever live. In other words, the problem is intergenerational in nature [36]. And given reasonable assumptions, the current population might be only a small percentage of all people to come. So, we will prefer to buck-pass. We will prefer to let future generations deal with the problem, should there be one. And here, unlike with merely global public goods, there are no obvious enforcement mechanisms. Future generations cannot even in principle control what we do. They do not yet exist.

[8] This is not to say that both models make the same predictions. Consider [10] Cleopatra case again. As Greene [19] points out, the social-preference model will not predict that your death by cancer at 25 could be justified because Cleopatra wanted an extra helping of dessert. For Cleopatra's *social* distance to someone living in 1021 is no different than her distance to someone in 2021. Given that the passing of time does not justify further discounting on the social-preference model, Cleopatra feels the same way toward the dessert-versus-death trade-off whether the unlucky soul who would die lives in 1021 or 2021 [19].

[9] For a discussion of discounting by political actors, see Ref. [33].

They cannot represent their interests. As such, the current generation has powerful self-interested incentives to pass the buck. And of course, as soon as one generation passes the buck, the same reasoning applies to every subsequent generation as it comes into being (cf. [17, pp. 403–404)].

So, if we see space colonization through the lens of existential risk mitigation, we should expect it to be subject to intergenerational buck-passing. We should expect, in other words, each generation to prefer collective action on space colonization be pursued to the next generation.

4. The importance of the impediments

We noted at the outset that discounting and buck-passing are not unique in their application to space colonization. So why exactly is there a special problem here?

4.1. Using co-benefits

It might be argued that there is no special problem. Any project that operates on long-time scales is challenging and—so the thought goes—space colonization is no different. We should simply refrain from trying to sell space colonization as a hedge against existential risk, even if that in fact is why we ultimately pursue it; rather, we should sell space colonization in terms of its short-term *co-benefits* and simply have colonization 'piggyback' off those efforts.

This idea is familiar from climate change mitigation (e.g., Hosking, Mudu, and Dora Ref. [24]), but Baum [3] suggests that the global catastrophic risk ('GCR') community writ large should take this lesson to heart and apply it to GCR-reducing activities in general (cf. [2]). For example, suppose we want a large-scale colony on the Moon. To sell it, we should emphasize the facilitation of helium-3 (^{3}He) extraction from the upper regolith. On Earth, our magnetic field and atmosphere provide protection from solar winds, making ^{3}He scarce. The moon lacks this protection. With nuclear fusion, ^{3}He is an abundant clean energy resource and a boon economically.[10] And then in due time, we get the colony.

The problem, however, is that large-scale ^{3}He extraction does not require space colonization. It only requires *space exploration*, and perhaps a commitment to a small manned lunar outpost, as seen in the 2009 film Moon, directed by Duncan Jones and starring Sam Rockwell. This falls well short of the kind of space colonization that would be required to be a meaningful hedge against GCRs—a point that will be substantiated further later. This is relevant, as it underscores that the gulf between a ^{3}He mining operation and space colonization appropriate for existential risk protection is quite large, both in terms of the money required and the further time it would take to complete a lunar colony once a mining based is established. If that is right, we can easily imagine the very same discounting trade-offs being operative even once the basic infrastructure for ^{3}He mining is in place. The benefit of being protected from existential risks will still accrue in the far future, and that must be weighed against more near-term benefits (like spending the money on building further mining infrastructure).

Of course, the feasibility of employing nuclear fusion and ^{3}He extraction on the moon is not a straightforward matter and not without its critics. But there is nothing special about this example. For the point is just this: it is doubtful that *any* commercial space interests on the radar require a colony or are made substantially easier by a colony. For instance, asteroid mining neither requires colonization nor (on current proposals) is made easier by a colony. More generally, absent a co-benefit that *requires* space colonization, or at least is made substantially easier by space colonization, the co-benefit strategy is inapt.

Note that none of this is to deny that there are no possible economic levers that might facilitate large-scale colonization. We might imagine, for instance, beginning with a very small-scale Martian colony and then leveraging the intellectual property developed along the way—technological developments spurred by the harsh reality of Martian life—into economic gains via licensing deals on Earth. That is not an improbable outcome and one with some analogs in history. But the question, then, is why we should think that these financial gains would then be re-invested to make a *much larger* space colony, especially when it is far from obvious that the rate of technological developments scales well with the size of the colony.

4.2. Time scales

Another complicating factor is the sheer time scales at which we are operating. This most obviously steepens the challenge presented by discounting, but it is reasonable to think that it likewise steepens the challenge presented by buck-passing. In principle, a public good can be minimally intergenerational merely by also benefiting the generation immediately subsequent to one's own. But that is not what is operative here; the full realization of space colonization, when designed to hedge against existential risks, will likely take many generations.

To see the import of this point, let us consider a comparison with climate change mitigation. It is well-known that action on climate change is hobbled by our discount rates. As Weisbach and Sunstein [49] note, the reason the recommendations of perhaps the two most famous policy reports on climate change—the 2007 Stern Report and the 2008 Nordhaus Report—diverge so radically is not that they disagree on which climate models are most likely to be correct, but because they employ wildly divergent discount rates, with Stern merely discounting by a catastrophe rate of 0.1%.[11] Consider too how discount rates impact the social cost of carbon (SCC), i.e., the measure (in dollars) of how much economic damage results from emitting 1 ton of CO2. The current model employed by the United States' Environmental Protection Agency—working with a discount rate of 3%—pegs the SCC at $51/ton. As Carleton and Greenstone [8] point out, at a 2% discount rate, the SCC goes up to $125/ton. The difference in SCC, driven by differences in discount rates, reflects an implicit position on what we think we owe future generations.

The fundamental ethical debate surrounding climate justice is often framed in terms of how we ought to allocate resources across generations: while the beneficiaries of (e.g.) our immediately cutting emissions will be people in future generations, the cost of these reductions will be paid by us, i.e., the current generation. This framing is not wrong, but it obscures something important. For while it is true that we (the current generation) will not suffer from the worst effects of climate change, we are suffering from its effects *now*. And while it is also true that beneficiaries of climate change mitigation will largely be future generations; these generations will likely be *near-future* generations—our children and our grandchildren. Yet despite this, we are *still* largely hobbled attempting to save face with cheap

[10] For a discussion of solar wind implanted ^{3}He distribution, see [28]. For nuclear fusion use and ^{3}He extraction technologies, see Simkos & Gray [46].

[11] The catastrophe rate is just the exogenous per year risk of humanity becoming extinct—in other words, the chance that the future benefits will not be realized.

talk and half-measures. This is again not to deny that there are exceptions—some countries do take climate change seriously. But it is noteworthy that these efforts do not come from the largest emitters on either an absolute or per capita basis.[12]

Matters are only worse when we compare this with the time scales involved in space colonization. It was noted at the outset that 'space colonization' here means a self-sustaining colony. Anything short of this is inapt for existential risk mitigation. But it cannot be *minimally* self-sustaining either, since that too would not be enough for existential risk mitigation: it would not help to have a colony which, while technically self-sustaining, is constantly on the edge of total collapse. So, let us consider as our target colony one large enough to sustain significant blows to its infrastructure and population while still serving its purpose of preserving humanity for the long term: a colony of one million people. Of course, all other things being equal, more would be better. Nevertheless, this number is useful as it approximates the population of a moderately large city, which is itself a reasonable approximation of the minimum amount of productivity required for a colony to be an effective (and so stable) hedge.[13]

There are many relevant factors when estimating how long it would take to establish a colony of this size. Since simplification cannot beg the question in favor of our thesis, we can focus on just one uncontroversial factor: food production. If a Martian colony relies on imports for food, it is *ipso facto* not self-sustaining. Thus, we can reframe the question as follows: how long would it take to develop a Martian colony that can produce its own food for one million colonists? In a recent study, Cannon and Britt [7] suggest one hundred years as a lower bound. This result assumes drastic shifts in dietary staples—one based on plants, insects, and cellular agriculture—and requires, amongst other things, nine thousand miles worth of vertically stacked twelve-foot-wide subterranean tunnels, lit with high power LEDs, and supplemented by sunlight collected from the surface.[14]

So, at one hundred years, we are looking at a minimum of five generations till we have a self-sustaining colony that is robust enough to serve as a hedge against existential risks. And given all various factors that we have left out—for example, simply the time it takes from now to safely develop a human-rated ship capable of taking colonists to Mars—it will likely take significantly longer than this. This does not bode well, given the psychological impediments in question. For since we cannot get our act together for benefits that will accrue for our own generation in the future, or for near-term generations, we should have little hope of doing so for benefits that accrue for generations at least one hundred years from now.

One might object that I am stacking the deck by requiring a colony of 1 million people: we can get by with far less for existential security, even with the assumption that certain production levels are required to stay off the knife's edge. Fine. Suppose I am wrong by half. Suppose that a 500,000-person colony would be sufficient. Even so, the temporal differences may not be significant: it might be that due to 'ramp-up' requirements regarding infrastructure etc., the time to go from 0 to 500,000 is vastly longer than going from 500,000 to 1 million. If so, then the especially deep impacts of discounting and buck-passing will still be operative.

4.3. The past as a guide to the future

It is tempting to consider discounting and buck-passing and still have hope for space colonization given our past actions. After all, mankind has undertaken many projects with long time horizons. Does not that provide reason to believe that space colonization will eventually come to be by collective means?

Take the Apollo Program, carried out by NASA and various industry partners from 1961 to 1972, as an example. The stated mission of the Apollo Program was not, of course, space colonization; it was to develop human space flight competency and, ultimately, lunar exploration. Still, it is illustrative insofar as it could be seen as a template for large-scale, long-term collective space-based project—a template that, if properly extended, might give us reason to think that the psychological impediments to space colonization can be overcome.

Unfortunately, the data do not bear this out. NASA's share of federal spending peaked in 1966 at 4.4% of federal spending. In 2018, NASA's share of federal spending was 0.5%.[15] On one estimate, direct R&D obligations for Apollo were $20.6 billion between 1960 and 1973. Adding "indirect" costs, such as the construction of facilities and operations, results in a $25.8 billion total cost. The Apollo Program would cost $264 billion in today's dollars. According to Dreier [12], "if the United States prioritized the project financially the same way it did in the 1960s, the nation would have to spend $641 billion to occupy the equivalent share of GDP." If we were to include the entire lunar effort—including both robotic missions and Gemini—these figures would rise to $288 billion and $702 billion, respectively.[16]

With these numbers in mind, consider now the public opinion. One survey indicates that in 1972—at the completion of the Apollo Program—over 60% of respondents felt that the United States was spending too much on space exploration (compared with less than 10% who felt we were spending too little). Decades later, with NASA budgets significantly lower than in 1972, over 30% of respondents still felt we were spending too much (compared with 20% who felt we were spending too little). Similar results have been obtained

[12] Just a few of many examples: In the United States, 12 states sued the Biden administration for reverting to an SCC of $51, where the Trump administration had set the SCC to as low as $1. See https://thehill.com/policy/energy-environment/542149-12-states-sue-biden-over-calculation-of-climate-impacts-in?rl=1. Senator Sen. Joe Manchin (D-WV) (https://www.vox.com/22709379/moderate-versus-progressive-democrats-climate)—an essential vote in the current US Congress' climate negotiations that are part of the Biden administration's "Build Back Better Plan"—has refused to back the Clean Electricity Performance Program (CEPP), which rewards energy suppliers who move away from fossil fuels. Experts argue (https://www.vox.com/22265119/biden-climate-change-renewable-energy-clean-electricity-standard-congress) that the CEPP is a critical component of any climate plan. At the recent Glasgow COP26 summit, while the resulting pact was promising in some ways—it included, amongst other things, codified rules for international carbon markets—it includes no enforcement mechanisms, with India insisting on "phase-down" as opposed to "phase-out" for fossil fuel use. This implies reduction, not elimination. We need the latter.

[13] One study [42] puts the minimal viable population at 110. However, the author admits that there are many risks in such a scenario that would threaten societal collapse, such as infertility, inbreeding, and random accidents that devastate production capability, both in manpower and critical machinery. Since we are considering space colonization as an x-risk mitigator, a self-sustaining colony for that purpose should be several orders of magnitude larger.

[14] One problem with surface-based greenhouses on Mars is that, although greenhouses are transparent, they will block an especially high portion of sunlight on Mars, since Martian greenhouses would require stronger material given the thin, cold air on Mars. Even on Earth, greenhouse windows only absorb 50 to 70% of light. See Ref. [7] for discussion.

[15] See https://usafacts.org/articles/50-years-after-apollo-11-moon-landing-heres-look-nasasbudget-throughout-its-history/.

[16] See Casey Drier in www.thespacereview.com/article/3737/1 for these figures in his work on a 'real accounting' for Apollo. For the first inflation adjustment, Drier uses NASA's New Start Index (NNSI). This adjustment answers the question, "How much would NASA be spending on Apollo today?" The second inflation adjustment (relative GDP share) answers, "If the US were to allocate resources to a space project to the same extent as it did for the lunar effort, how much would NASA be spending today?" See also Ref. [34].

across multiple surveys.[17] The point is not that we do not want to go to space. It is that we do not want to pay for it.

Putting these points together with the geopolitical context in which the Apollo Program was conducted, we see that it is an overly rosy picture to assume that the Apollo Program was really about science or exploration, even if these were the stated aims. Rather, it was to defeat the USSR, symbolized in terms of beating them to the moon.[18] And while there were no doubt other factors, once that aim was achieved in 1969, interests in further space pursuits waned ([37]; cf. [32]). This is evidenced by both the precipitous drop in funding and public opinion.

A general lesson emerges here. The USSR was not merely a hypothetical threat or a threat that the United States had no reason to be worried about anytime soon. Immersed in the Cold War, the USSR achieved the first successful satellite launch with Sputnik in 1957 and put the first man into space four years later. This made aerospace technology not just a symbol of national pride, but also military power, with President Kennedy fearing relative disparities in missile technology.[19] It is unsurprising then that this was enough to motivate the United States into investing into the Apollo Program. Yet it is also hardly ideal. In fact, inferring anything from the Apollo Project about our appetite for long-term investments on x-risk mitigation is misguided. Imagine that we unexpectedly find out that an asteroid is going to hit us in three years, and when it does, it will destroy all life on Earth. It would be no surprise if, upon hearing this news, there was a massive collective effort toward space colonization. Alas, this is cold comfort. For the hope of those who have rang the alarm on x-risks is that the mitigation efforts are done *before* it is too late—that is, before the threats are knocking on our doorstep, so to speak. If an asteroid hits Earth, we should watch from a fully colonized Mars, not while in transit from the first colonization ships.

5. S-risks and institutional reform

The claim I hope to have established thus far is this: as things stand, efforts toward large-scale, self-sufficient, x-risk−reducing space colonies by broadly democratically constituted nations face steep challenges from our general tendencies to discount the future and pass the buck to subsequent generations. I want to close by exploring where this leaves us. I will make two points. The first concerns the prospects for space colonization when discounting and buck-passing are not operative. The second point—made in light of the first—concerns policy and institutional reform recommendations to help ensure not only that we colonize space, but that space colonization also goes well.

We defined an x-risk as any risk that threatens the destruction of our long-term potential, either through literal extinction or an unrecoverable collapse. Now consider *suffering risks* ('s-risks'): risks of events that bring about suffering in astronomically significant amounts. Some s-risks might count as x-risks (e.g., a totalitarian regime that both includes immense amounts of suffering and precludes the realization our potential), but it is clear that not all x-risks are s-risks (e.g., a scenario in which the universe is instantly emptied of all sentient beings). In any event, the notion of an s-risk is intuitive enough to see why it is at least *prima facie* plausible that space colonization might have catastrophically negative outcomes.

There are several reasons for this. One idea from Torres [47, p. 79] is that by expanding into space, we will eventually see drastic levels of phylogenetic and ideological diversity, resulting in

> civilizations [with] entirely different normative preferences, moral tendencies, and even scientific institutions. Some will almost certainly be violence-inclined, thus giving others an incentive to strike first. Even more, diversity with respect to cognition, emotionality, and language will undercut the mutual trust needed for otherwise irenic civilizations to avoid spirals of militarization or defect in the prisoner's dilemma predicaments. Thus, a colonized cosmos would be an arena poised and spring-loaded for violence.

What is more, many of the standard ways of preventing widespread, interminable conflict—like creating a 'cosmic' analog of Hobbes' Leviathan—might very well be unworkable [47]. It might be unworkable simply because of the vastness of space itself, making communication and collective action practically impossible.[20]

With this point in place, we can carve out two sources of s-risks. Consider the following first:

> SLOPPY IMPLEMENTATION: S-risks that are sourced to sloppy implementation of space colonization—i.e., space colonization done under duress, as opposed to being done well in advance of an existential crisis.

The idea behind SLOPPY IMPLEMENTATION is straightforward: we are more liable to make mistakes—potentially uncorrectable mistakes—if we try to pull off a large-scale project when under duress. Suppose—to use our earlier example again—we learn that an asteroid is going to strike Earth in three years. Alongside asteroid deflection measures, the major nations of Earth come together to build a large-scale, self-sufficient colony on Mars as a backup to ensure humanity survives in case the deflection measures fail. Attempting to complete such a project in three years—no matter how much resources are thrown at it—is a recipe for disaster. Yet given our tendencies toward high discounting and buck-passing, it should be of no surprise that large-scale, self-sufficient space colonization happens in precisely this manner, since under these conditions, the force of these tendencies would no longer be operative. After all, we would be saving ourselves, not our far future brethren.

The second source is as follows:

> BAD CORRELATIONS: S-risks that are sourced to 'bad correlations' between power structures that instantiate one quality that is good for space colonization (e.g., low discounting) and instantiating other qualities that are bad for s-risks.

On BAD CORRELATIONS, it is important to realize that features which are good for space colonization like low discounting do not necessarily correlate with features that are portend low s-risks. Here is one example. We know that liberty and the realities of survival in space are in tension to some degree (e.g., Refs. [9,35,39,45]). Yet we might imagine a society—perhaps a 'stable' totalitarian society—whose (sole) ruling party is obsessed with locking in their dominance for the long term and thus possesses low discount rates insofar as it puts a major portion of its resources to that end. This society might pursue space colonization, and be incredibly efficient at it, as totalitarian societies tend to be effective at mobilizing resource by

[17] See the National Opinion Research Center's General Social Survey on public opinion about spending on space exploration. While the percentage of people who believe NASA spending is too high has dropped significantly since the immediate post-Apollo period, as of 2020, almost 50% believe that NASA spending is "about right". It is not unreasonable to assume that this reflects people's priorities, which is concerning given that current NASA spending is not even remotely enough to sustain large-scale colonization efforts. Indeed, it is not enough to sustain much smaller scale projects, like a return to the moon.

[18] See Ref. [48] for an extensive treatment of this and other related issues.

[19] See Ref. [40] for discussion.

[20] For a detailed analysis of the risks (s-risks and otherwise) of space colonization, see Ref. [13].

brute force [31]. This is no happy result, however, since a totalitarian society aimed at mitigating extinction risks at all costs will, amongst many other things, likely exacerbate the lack of freedom already inherent in space colonization. Depending on the scope of the society, this could cause misery at unforeseen levels and thus constitute a suffering catastrophe.[21]

We do not, then, want to only pursue space colonization under duress or via power structures that would curtail human flourishing. And it is quite unlikely that we can, in the short term, change human psychology writ large in a way that favors safeguarding the future. But what we *can* do—and what I want to suggest here—is implement institutional reforms that *work around* our general biases against the future. Following John and MacAskill [27], we can think of these reforms as *longtermist* institutional reforms. This is not necessarily longtermist in the strong sense of implying that what matters most about our actions is their far future actions.[22] Rather, these reforms are longtermist insofar as they start from a recognition of the deep moral importance of safeguarding civilization—not just by ensuring its continued long-term existence, but also by ensuring we realize our potential, whatever it may be.

The structure of large, democratic governments as they are currently constituted is by and large antithetical to the longtermist ethos. We noted that election incentives are one institutional determinant of this: performing well in 2- to 5-year election cycles requires prioritizing policies with short-term payoffs and (even if the politician herself favors lower discount rates) pandering to the myopia of one's constituency. Having longer election cycles, however, will not fix the problem on its own—for one thing, it would not fix the underlying psychological biases that generate the short-termist behaviors (like discounting) in political actors and constituencies that exhibit them—and long elections cycles are not unproblematic on their own terms. Instead, as John and MacAskill [27] suggest, we could create a novel representative deliberative body whose sole mandate is to represent the interests of future generations, working in an advisory capacity: *futures assemblies*. The ‘sole mandate’ provision would help ensure that members would be focused on the long term, and if modeled on regular citizens' assemblies and thus not dependent on elections and fundraising, futures assemblies would not be hamstrung by the short-term incentives that paralyze other political institutions.[23]

A more radical proposal suggested by John and MacAskill [27] is to create legislative houses dedicated toward the protection of future generations, which would work together with lower houses dedicated toward present-focused legislation. As John and MacAskill [27] note, we could envision legislation requiring passage in both houses to become law. In this way, each house—one directed toward current generations and one directed toward future generations—could act as a check on the other. This would help ensure that present-focused concerns are not ignored for future-focused concerns, and vice versa.

[21] Schmidt and Bohacek [43] illustrate the many ways in which the relationship between systems of governance and the creation and flourishing of space colonies is complex, advocating the incorporation of cosmopolitanism. One issue here that is especially pertinent to space colonies built primary to reduce x-risks is the countervailing pressures between epistemic authorities (whose technical expertise is needed to ensure the health and security of the colony's infrastructure) and long-term questions of political legitimacy, once the colony is on surer footing.

[22] For discussion of this strong axiological version of longtermism, along with its deontological cousin, see Ref. [18].

[23] Sweden has something like this, although it seems as though it was only a temporary basis. See https://www.government.se/government-of-sweden/prime-ministers-office/mission-the-future/. In fiction, this idea can be found in Kim Stanley Robinson's climate change novel “The Ministry for the Future”.

These reforms would not, of course, be implemented simply for the sake of colonizing space. But insofar as one thinks that space colonization is essential for safeguarding civilization, these reforms would help us not only colonize space, but also ensure that when we do, we do it in a way that lowers s-risks. They would do this by countering the psychological obstacles that make myopia the unfortunate norm.[24]

Author statement

All aspects of the manuscript were completed by its sole author (Joseph Gottlieb). This includes:

- Conceptualization
- Formal analysis
- Methodology
- Writing – original draft, review, and editing.

Declaration of competing interest

The author declares that he has no conflicting interests.

References

[1] M. Basner, et al., Psychological and behavioral changes during confinement in a 520-day simulated interplanetary mission to Mars, PLoS One (2014), https://doi.org/10.1371/journal.pone.0093298.

[2] S. Baum, D. Denkenberger, J. Haqq-Misra, Isolated refuges for surviving global catastrophes, Futures 72 (2015) 45–56.

[3] S. Baum, The far future argument for confronting catastrophic threats to humanity: practical significance and alternatives, Futures 72 (2015) 82–96.

[4] S. Baum, T. Barrett, Towards an integrated assessment of global catastrophic risk, in: B.J. Garrick (Ed.), Proceedings of the First International Colloquium on Catastrophic and Existential Risk, Garrick Institute for the Risk Sciences, University of California, Los Angeles, 2018, pp. 41–62.

[5] N. Beckstead, On the Overwhelming Importance of Shaping the Far Future (PhD thesis), Rutgers University, 2013. Available at: https://rucore.libraries.rutgers.edu/rutgers-lib/40469/PDF/1/play/.

[6] N. Bostrom, Existential risk prevention as global priority, Glob. Policy 4 (1) (2013) 15–31.

[7] K.M. Cannon, D.T. Britt, Feeding one million people on Mars, New Space 7 (4) (2019), https://doi.org/10.1089/space.2019.0018.

[8] T. Carleton, M. Greenstone, Updating the United States Government's Social Cost of Carbon, University of Chicago, Becker Friedman Institute for Economics, 2021, https://doi.org/10.2139/ssrn.3764255. Working Paper No. 2021-04, Available at SSRN: https://ssrn.com/abstract=3764255.

[9] C. Cockell (Ed.), Human Governance beyond Earth: Implications for Freedom, Springer, 2015.

[10] T. Cowen, D. Parfit, Against the social discount rate, in: P. Laslett, J.S. Fishkin (Eds.), Justice between Age Groups and Generations: Philosophy, Politics, and Society Sixth Series, Yale University Press, New Haven and London, 1992, pp. 144–161.

[11] M.L. Cropper, S.K. Aydede, P.R. Portney, Rates of time preference for saving lives, Am. Econ. Rev. 82 (1992) 469–472.

[12] C. Dreier, A New Accounting for Apollo: How Much Did it Really Cost? The Space Review, 2019. https://www.thespacereview.com/article/3737/1.

[13] D. Deudney, Dark Skies: Space Expansionism, Planetary Geopolitics, and the Ends of Humanity, OUP, Oxford, 2020.

[14] E. Finneron-Burns, What's wrong with human extinction? Can. J. Philos. 47 (2–3) (2017) 327–343, https://doi.org/10.1080/00455091.2016.1278150.

[15] S. Frederick, Measuring intergenerational time preference: are future lives valued less? J. Risk Uncertain. 26 (2003) 39–53.

[16] J. Frick, On the survival of humanity, Can. J. Philos. 47 (2–3) (2017) 344–367, https://doi.org/10.1080/00455091.2017.1301764.

[17] S.M. Gardiner, The real tragedy of the commons, Philos. Publ. Aff. 30 (4) (2001) 387–416.

[18] H. Greaves, W. MacAskill, The case for strong longtermism, GPI Working Paper No. 5-2021, https://globalprioritiesinstitute.org/hilary-greaves-william-macaskill-the-case-for-strong-longtermism-2/, 2021.

[19] P. Greene, M. Sullivan, Against time bias, Ethics 125 (4) (2015) 947–970.

[24] For helpful comments and discussion, I wish to thank the audience at the SSoCIA 2020 meeting, the SWIG working group, Seth Baum, James Schwartz, Kelly Smith, Bruce Tonn, and two anonymous referees for Space Policy.

[20] Greene, P. (2022) Social Bias and Not Time Bias. (Unpublished manuscript). https://static1.squarespace.com/static/56d45b30f699bb6f0beeaf5a/t/6219eff17e85c74fcdfb9ef3/1645866994345/Social+Bias+and+Not+Time+Bias.pdf.
[21] J. Gottlieb, Space colonization and existential risk, J. Am. Philos. Assoc. 5 (3) (2019) 306–320.
[22] J. Haqq-Misra, Can deep altruism sustain space settlement? in: K. Szocik (Ed.), The Human Factor in a Mission to Mars: an Interdisciplinary Approach Springer, 2019, pp. 145–155.
[23] O.P. Hauser, D.G. Rand, A. Peysakhovich, M.A. Nowak, Cooperating with the future, Nature 511 (2014) 220–223.
[24] J. Hosking, P. Mudu, C. Dora, Health in the Green Economy: Health Co-benefits of Climate Change Mitigation—Transport Sector, World Health Organization, Geneva, 2011.
[25] J. Jacquet, et al., Intra- and intergenerational discounting in the climate game, Nat. Clim. Chang. 3 (2013).
[26] Magnus Johannesson, Per-Olov Johansson, Saving lives in the present versus saving lives in the future—is There a framing effect? J. Risk Uncertain. 15 (1997) 167–176.
[27] T. John, W. MacAskill, Longtermist institutional reform, in: N. Cargill, T. John (Eds.), The Long View, FIRST, London, UK, 2021, pp. 44–60.
[28] J.R. Johnson, T.D. Swindle, P.G. Lucey, Estimated solar wind-implanted helium-3 distribution on the Moon, Geophys. Res. Lett. 26 (3) (1999) 385–388.
[29] H. Jones, The recent large reduction in space launch cost, in: Information International Conference on Environmental Systems (ICES), 2018. https://ntrs.nasa.gov/citations/20200001093.
[30] N. Kanas, Humans in Space: the Psychological Hurdles, Springer, 2015.
[31] B. Kaplan, The totalitarian threat, in: Cirkovic Bostrom, Rees (Eds.), Global Catastrophic Risks, Oxford University Press, Oxford, 2008.
[32] M.J. Logsdon, Space in the post-cold war environment, in: S.J. Dick, R.D. Launius (Eds.), Societal impact of Spaceflight, NASA History Division, Office of External Relation, Washington, 2007, pp. 90–102.
[33] M.K. MacKenzie, Institutional design and sources of short-termism, in: Gonzalez-Ricoy, Gosseries (Eds.), ´ Institutions for future generations, Oxford University Press, UK, 2016, pp. 24–48.
[34] A.C. McDonald, The Long Space Age: the Economic Origins of Space Exploration from Colonial America to the Cold War, Yale University Press, 2017.
[35] T. Milligan, Rawlsian deliberation about space settlement, in: C. Cockell (Ed.), Human Governance beyond Earth. Space and Society, Springer, Cham, 2015.
[36] T. Ord, The Precipice: Existential Risk and the Future of Humanity, Hachette Books, 2020.
[37] E. Panagiotarakou, Agonal conflict and space exploration, in: T. Milligan, J. Schwartz (Eds.), The Ethics of Space Exploration, Springer, 2016.
[38] D. Parfit, Reasons and Persons, Oxford University Press, Oxford, 1984.
[39] E. Persson, Citizens of Mars Ltd, in: C. Cockell (Ed.), Human Governance beyond Earth. Space and Society, Springer, Cham, 2015.
[40] C.A. Prebble, “Who ever believed in the ‘missile gap’?”: John F. Kennedy and the politics of national security, Pres. Stud. Q. 33 (4) (2003) 801–826.
[41] F.P. Ramsey, A mathematical theory of saving, Econ. J. 38 (152) (1928) 543–559.
[42] J. Salotti, Minimum number of settlers for survival on another planet, Sci. Rep. 10 (2020) 9700, https://doi.org/10.1038/s41598-020-66740-0.
[43] N. Schmidt, P. Bohacek, First space colony: what political system could we expect? Space Pol. 56 (2021) 101426.
[44] S. Schubert, L. Caviola, N.S. Faber, The psychology of existential risk: moral judgments about human extinction, Sci. Rep. 9 (2019) 15100, https://doi.org/10.1038/s41598-019-50145-9.
[45] J. Schwartz, Lunar labor relations, in: C. Cockell (Ed.), Dissent, Revolution, and Liberty in Space, Springer, 2016, pp. 41–58.
[46] T. Simkos, M. Gray, Lunar helium-3 fuel for nuclear fusion: technology, economics, and resources, World Futures Rev. 6 (2) (2014) 158171, https://doi.org/10.1177/1946756714536142.
[47] P. Torres, Space colonization and suffering risks: reassessing the “maxipok rule”, Futures 100 (2018) 74–85.
[48] N. Tyson, A. Lang, Accessory to War: the Unspoken Alliance between Astrophysics and the Military, W. W. Norton & Company, 2018.
[49] D. Weisbach, Cass R. Sunstein, Climate Change and Discounting the Future: A Guide for the Perplexed, vol. 433, Yale Law and Policy Review, 2009.