Comparative Axiological Assessment of Space Colonisation

Caleb Peppiatt - CRS Essay Draft

Contents:

Introduction

Under various theories of value, the decision to initiate irreversible space colonisation activities might be one of the most significant that humanity could make. The value of these activities diverges starkly under different perspectives of what is valuable. As such, it is important to first delineate which theories of value favour space colonisation, and which recommend refraining, and second, to assess the overall desirability of space colonisation given moral uncertainty.

By "space colonisation activities," this essay is chiefly interested in the following processes:

Category:	Description:	Examples:
A	Actions that initiate self-perpetuating processes which propagate without further authorisation.	Directed Panspermia. Von Neumann Probes. Establishing colonies capable of autonomously launching further expansion.
B	Actions that durably entrench particular values or governance mechanisms.	Creating singletons, such as those governed by ASI. AI-enabled surveillance and enforcement systems.
C	Creating populations that cannot be feasibly governed at scale.	Democratising access to synthetic phenomenology, terraforming, and radical genetic engineering. Establishing economies reliant on sentient digital workers and emulated minds.

Two Category A activities - directed panspermia and establishing autonomous colonies - shall be analysed in depth as they are relatively easy to assess under each normative position. The following "comparative axiological assessment" compares a select sample of ten positions, which were chosen due to their impact on debates in space ethics. Positions vary by theory of value (what is deemed intrinsically valuable, as in Owe, Baum, and Coeckelbergh, 2022), by population ethics (how they weigh generations), and by normative commitments (how the choice-worthiness of decisions are assessed).[1]

[1] Anthropocentric perspectives that do not consider non-human entities as holding moral significance shall not be considered. If the reader is an anthropocentrist with regards to moral standing, they may interpret position (9) *Rawlsian contractarianism* in anthropocentric terms.

The first four positions share normative stances - that the choiceworthiness of a decision depends upon its expected consequences, and that intrinsically valuable aspects of reality ought to be promoted or maximised. They further agree that valenced phenomenal experiences are what is valuable and disvaluable. However, they vary considerably with regards to population ethics - how to weigh different generations, and axiological symmetry - how to weigh positive and negative experiences against each other.

(1) Total utilitarianism. This position is sentientist (positively valenced experience is intrinsically valuable), symmetric (positive experiences are valuable equivalent to how negative experiences are disvaluable), impartial (the value of positively valenced experience does not depend upon species-membership or location in time and space), totalist (the aggregate value of a state of affairs is relevant for assessing desirability), and consequentialist (a decision's choiceworthiness depends upon its consequences). This is the position generally held in the longtermist case for space colonisation (Greaves and MacAskill, 2021).

(2) Moderate person-affecting utilitarianism. This position is similarly sentientist, symmetric, and consequentialist, yet crucially weighs currently-existing persons as being of greater moral significance than potential future persons (Bader, 2022).[2]

(3) Moderate suffering-focused utilitarianism. This position is sentientist, ideologically asymmetric (negatively-valenced experiences, i.e. suffering, is of greater disvalue than positively-valenced experiences are valuable), impartial, and consequentialist. This is the position held in the author's critique of the longtermist case (Peppiatt, 2026).[3]

(4) Negative utilitarianism. This position is sentientist yet entirely asymmetric (negatively valenced experience, i.e. suffering, is intrinsically disvaluable, yet there are no equivalent "positive" counterparts). It is consequentialist, as a decision's choiceworthiness depends upon the extent to which it reduces, alleviates, or prevents suffering. (For a nuanced proponent, see Leighton, 2023).

The following three positions vary considerably in terms of what they posit to be intrinsically valuable. For simplicity and ease of comparison, they shall be interpreted as consequentialist, and that intrinsically valuable aspects of reality ought to be maximised rather than merely conserved (Baum, 2016, 5).

[2] This position is arguably supported by intuitions in favour of the procreative asymmetry, non-identity problems, and arguments that an impersonal "God's-eye-view" of potential persons is objectionable.

[3] This position is arguably supported by intuitions regarding the phenomenological intensity of severe suffering, its felt badness, and the urgency to relieve it.

(5) Biocentric individualism. This position is biocentric - living organisms are intrinsically valuable, regardless of whether or not they are sentient. Originally defended by environmental ethicists Paul Taylor and Robin Attfield, it has substantially informed debates in space ethics (Brennan and Lo, last update 2021).

(6) Ecocentric holism. This position is ecocentric - whole biospheres are valuable as a collective, rather than individual organisms therein. The Earth's whole life-story is what has intrinsic value, as well as holistic entities such as species and ecosystems, such that it ought to be expanded into other star systems via constructing other environments in which Earth's biosphere can thrive (Owe, 2023, 2).[4]

(7) Nonhumanist perfectionism. Nonhumanist perfectionism seeks to promote objectively good states of affairs that may be independent of what is good for humans (Steven, last update 2021). Rather than simply welfare, perfectionists value a range of impersonal objective goods, such as scientific discovery, technological achievement, artistic greatness, and natural beauty. Arguably, many longtermists advocating for space colonisation implicitly appeal to perfectionist intuitions when they speak of the "cosmic endowment."

Three further positions are included as they forward normative stances distinct from the consequentialist impulse to promote certain goods.

(8) Deontological constraints. This position posits several ethical constraints that constrain the choiceworthiness of certain decisions even if they were to lead to greater expected value. Three particularly suggestive constraints can be outlined: Parfit's Consent Principle holds that it is wrong to treat anyone in any way to which that person could not rationally consent (Parfit, 1984); that our choices regarding the distribution of resources and capabilities ought to be justifiable to future individuals ("Deontic Egalitarianism" per Meyer, 2018); that future individuals should not be subjected to domination (following the republican conception of freedom in Pettit, 1999).

(9) Rawlsian contractarianism. This position contends that just institutional arrangements are those which rational parties would agree to under fair conditions, such as a veil of ignorance wherein parties know their higher-order interests yet not particular facts about themselves (Ligor and Matthews, 2022).[5]

[4] This position is arguably supported by intuitions that diverse, complex biospheres are intrinsically valuable, that the extinction of species is tragic above and beyond the tragedy of the deaths of individual organisms, and that it would be good to preserve life if Earth-like biospheres are cosmically rare.

[5] Note that Rawls does not directly address the identity of future generations, and so it is not particularly clear how Rawls' principles apply to identity-affecting choices.

(10) Virtue ethics. This position holds that the choiceworthiness of an action in part depends on whether the action exemplifies virtues and produces a virtuous character. In debates in mainstream space ethics, a virtue ethical approach is promoted by McKay's 1990 environmental ethics wherein human actions ought to exemplify humility (rather than hubristic interventionism and the domination of an environment), wise stewardship (respecting intergenerational sustainability), and relations of care.[6]

These ten positions are not intended to be exhaustive or comprehensive in their coverage. Instead, they have been chosen as they vary in their theories of value, perspectives on population ethics, and how they normatively assess the choiceworthiness of actions. The following section will show that the desirability of space colonisation activities considerably varies depending on which position is assumed.

[6] I am here interpreting "virtue ethics" broadly so as to include Aristotle's Nicomachean Ethics (defended more recently in MacIntyre, 1981), while specifying McKay's environmental ethics as one relevant application.

Assessing Colonisation

Having outlined ten positions by which space colonisation activities can be assessed, this section shall now apply them. First, let us consider the category (A) action of directed panspermia, or seeding life on other planets.[7]

- Under total utilitarianism, seeding life may be highly disvaluable given the aggregate future suffering experienced by sentient organisms. The sign of this verdict crucially depends upon whether the aggregate experience of wild animals is net-negative, such that resolving this uncertainty prior to seeding life is recommended (Sivula, 2022, 180, 193; Sivula, 2025).
- Under a moderate person-affecting utilitarianism, the moral significance of wild animal populations that would potentially exist in millions or billions of years is of much lesser weight than currently-existing sentient beings. This suggests that the value or disvalue of directed panspermia is of less monumental significance than under total utilitarianism.
- Under a moderate suffering-focused utilitarianism - which puts greater weight on the suffering instantiated by evolutionary processes relative to symmetric utilitarianism - refraining from panspermia is plausibly recommended (Peppiatt, 2026, 12-13, 19-21).
- Under negative utilitarianism, panspermia is strongly disfavoured, as it would create vast suffering populations lasting for potentially billions of years- with no counterbalancing upside.
- Under biocentric individualism, panspermia is plausibly recommended due to the vast number of biological organisms that could be seeded (Mautner, 2000). If the biocentrist is concerned about spreading life - irrespective of the suffering experienced by living sentient organisms - then the only significant cause for concern might be the displacement of native microorganisms due to forward contamination (Lupisella, 1997).
- Under ecocentric holism, panspermia may be a crucial duty for humanity (Owe, 2023, 2; Gros, 2016, 12). After all, if the diversity and flourishing of biospheres is what is intrinsically valuable, rather than individual organisms, then wild animal suffering is ultimately a valuable incidental aspect of maintaining a diverse biosphere. Sources of concern would instead relate to sterilising biospheres via contamination incidents, as well as potential conflicts with extra-terrestrial civilisations (Baum, 2016, 11). There is also peer disagreement among ecocentrists as to whether artificially seeded biospheres are equally valuable to ones which naturally arose.
- Under nonhumanist perfectionism, panspermia may be valuable or disvaluable depending on the objective goods posited. For instance, if unique abiotic landscapes are valuable, each with distinctive world-histories, then panspermia may be disvaluable insofar as it disrupts the integrity of other planets (Rolston, 1986). Alternatively, *if* goods such as biological diversity, relations of care, or instances of courage are deemed to be intrinsically valuable, *and if* such goods outweigh disvaluable states, then seeding evolutionary processes may be highly choiceworthy.
- Given deontological constraints, panspermia may be disfavoured: sentient organisms might not be able to rationally consent to being subject to evolutionary processes; the distribution of capabilities

[7] The following descriptions may apply *mutatis mutandis* to terraforming and other processes of biosphere creation.

or conditions for welfare are not justifiable for the majority of species; and being born into an isolated and unsupervised biosphere amidst hostile environments and conditions of scarcity and antagonism might be said to constitute domination (see Peppiatt, 2025, 65-76).

- Given Rawlsian contractarianism, the crucial consideration is whether "rational parties" behind the veil of ignorance are limited to humans (Rowlands, 2002). If non-human animals are excluded, and if panspermia has minimal impact on questions of distributive justice on Earth, then panspermia may not be strongly favoured or disfavoured.
- Finally, virtue ethics might plausibly view irreversible interventions into other star systems as hubristically "playing God," or else as a courageous effort to preserve life in the cosmos (for a discussion of whether the decision to seed evolutionary processes is virtuous even for a deity, see Peppiatt, 2025, 80).

Second, let us consider the category (A) action of establishing autonomous colonies.

- Under total utilitarianism, the value of establishing such colonies is somewhat indeterminate. Given non-negligible risks of astronomical suffering associated with farming practices, totalitarianism, and warfare in space, and given that technologies enabling suffering to be instantiated at scale are relatively feasible compared with suffering-mitigating technologies, seeding such colonies may be highly disvaluable (Peppiatt, 2026, 24-33, 45-54). Restraint is advisable given widely-used precautionary principles (Peppiatt, 2026, 65-7).
- Under a moderate person-affecting utilitarianism, the moral standing of potential future populations on seeded colonies carries less weight compared to that of currently-existing populations. It is possible that attempting to establish independent colonies would harm the welfare of currently-existing terrestrial populations given climate crises and how activities in space alter the stability of geopolitics (Deudney, 2020; Kemp, 2025).
- Under a moderate suffering-focused utilitarianism, there is further reason to refrain from such activities, as the disvalue of severe suffering risked by colonising space is weighed more heavily compared to equivalent positive experiences.
- Under negative utilitarianism, establishing independent colonies capable of further expansion would be an awful decision, as it could potentially bring into existence vast numbers of future lives - and by extension, suffering. One caveat is whether colonisation could enable humanity to prevent and reduce suffering of extraterrestrial organisms (Peppiatt, 2026, 55-62).
- A biocentric individualist might favour establishing colonies if human colonists were able to responsibly terraform planets and establish biospheres containing large and diverse biological populations. This state of affairs might be considered less valuable if organisms were merely instrumentalised as food sources (as plausibly would be the case for both surface and BLSS farming, Peppiatt, 2026, 17-18).
- An ecocentric holist similarly might favour establishing colonies if human colonists could responsibly terraform planets and establish biospheres. This might partly depend on whether such biospheres were sufficiently "natural" and free from excessive human intervention (Katz, 1991 and 1997), and whether Earthlike diversity and creativity can flourish in non-terrestrial environments.

- Under nonhumanist perfectionism, establishing such colonies might be highly valuable depending upon the specific intrinsic values: colonising space would provide many sources of aesthetic appreciation, scientific discovery, technological and artistic achievement, and novel societal forms, even if the integrity of natural environments is undermined.
- Deontological constraints plausibly cut against the decision to establish other colonies. For instance, human transit to other star systems might require generation ships which future generations could not rationally consent to being born into, as well as conditions of political domination under totalitarian regimes with minimal exit options (Peppiatt, 2026, 25-7, 34-5).
- Rawlsian contractarianism applies most cleanly to near-term, anthropocentric distributive contexts. The likely socioeconomic composition of isolated space colonies is an open question, yet given resource extraction in space might result in extreme wealth inequalities, the veil of ignorance might suggest that we ought to only establish independent colonies once certain norms and regulations concerning property rights have been established.
- Finally, the decision to establish independent colonies might exemplify certain virtues such as courage (although note the genealogical critique of Becker, 2025, c. 5), and forming societies in hostile environments under conditions on non-domination requires an impressive degree of solidarity (such as Le Guin's Anarres). Others might contend that the acquisitive drive to expand to new frontiers neglects the act of cherishing our current planet, and that the virtues associated with scientific discovery can still be exemplified within the confines of our own solar system.

Comparative Matrix

For each position, the following matrix summarises the "sign" - whether the action is favoured or disfavoured relative to restraint - and relative conditions underlying axiological assessments.

Position	Directed Panspermia	Establishing Independent Colonies
Total utilitarianism	Sign is indeterminate, depends on whether aggregate wild-animal welfare is net-positive or net-negative.	Sign is indeterminate, yet precautionary restraint is favoured if suffering-enabling technologies are relatively feasible.
Moderate person-affecting utilitarianism	Sign is indeterminate, depending on animal welfare as above, yet stakes are less severe.	Plausibly negative if colonisation harms current populations (climate crises, geopolitical instability).
Moderate suffering-focused utilitarianism	Refraining is favoured more strongly than under symmetric views, depending on animal welfare as above.	Refraining is favoured more strongly than under symmetric views.
Negative utilitarianism	Sign is strongly negative; refraining is recommended.	Sign is negative and increases with population size, yet also depends upon whether humanity alleviates ET suffering.
Biocentric individualism	Sign is positive if it increases the number of living organisms; it also depends on whether native microbial life is displaced.	Sign is positive if colonists create large, diverse, and responsibly-terraformed biospheres.
Ecocentric holism	Sign is potentially highly positive; distant biospheres must then be protected from sterilisation.	Sign is positive if colonists create biospheres that are sufficiently natural and diverse.
Nonhumanist perfectionism	Sign depends on which objective goods are posited. Disvaluable if unique abiotic landscapes have standing.	Sign is likely positive, conditional on certain aesthetic and epistemic goods being sufficiently realised.
Deontological constraints	Disfavoured given constraints of consent, just capability distributions, and	Disfavoured for cases such as generation ships due to domination and inability to

	non-domination.	consent to conditions.
Rawlsian contractarianism	Sign is indeterminate; depends on whether non-human parties are admitted, and effects on terrestrial distributive justice.	Conditionally disfavoured until adequate property and governance norms are established.
Virtue ethics	Sign depends on which virtues are prioritised and how the action is framed.	Under norms of terrestrial stewardship, colonising might be disfavoured; likewise depends upon selection of virtues.

Discussion

The comparative axiological matrix suggests that assessing the value of particular space colonisation activities depends significantly on one's axiological positions. For instance, a negative utilitarian might conclude that Category A space colonisation activities are highly undesirable, given that they could seed large suffering populations, while a biocentric individualist might conclude that such activities are highly desirable given that living organisms might proliferate across the cosmos. How might we reason given that value assessments diverge so starkly between views?

Suppose that one's preferred normative position either strongly favours or strongly disfavours space colonisation. Given moral uncertainty, an initial, intuitive response might be to moderate one's assessment in the high desirability or high undesirability of colonisation. After all, if the assessment of desirability varies considerably between positions, and if one has a non-negligible credence that other positions are correct, then intuitively one ought to moderate away from tail-end assessments. While the extent to which one's credence in the desirability of space colonisation ought to be moderated cannot be precisely quantified, the existence of contradictory desirability assessments may suggest that one's credence ought to be moderated to a significant degree. However, this assumes that each normative position is comparable, and that an overall assessment of the desirability of space colonisation can be made. Is it possible to produce an overall assessment of desirability under conditions of moral uncertainty, wherein we are uncertain which normative position is correct?

The first option is to contend that several of these positions *cannot* be intertheoretically compared, such that an overall assessment of desirability cannot be undertaken. It may be the case that positions as varied as

anthropocentric Rawlsian contractarianism, moderate person-affecting utilitarianism, and ecocentric holism lack comparable theoretic units or ways to assess the choiceworthiness of actions. This itself would not necessarily paralyse decision-making, as one might first establish conditional conclusions (e.g. "if we assume biocentric individualism..."), distinguish between sources of normative uncertainty and of empirical uncertainty, and work to minimise the latter while explicitly acknowledging the former.

Several instances of apparently normative disagreement might instead be instances of unresolved empirical questions: the overall expected value sign for total, moderately suffering-focused, and negative utilitarian positions might depend upon the valence of wild-animal welfare, and whether particular risks of astronomical suffering can be reliably avoided. Similarly, biocentric, ecocentric, and virtue-ethical assessments might each depend on whether space colonisation activities can be responsibly governed, such that native organisms, seeded biospheres, and human societies are each respected. Rawlsian contractarian, moderate person-affecting utilitarian, and deontological assessments might plausibly depend upon the governing institutions and norms affecting near-term colonisation initiatives and their distributive effects on terrestrial societies and human colonists. While many normative commitments are irreducible and potentially incommensurable, several positions might agree on whether colonisation is desirable or not depending on underlying empirical uncertainties. This finding is robust regardless of whether overall intertheoretic assessments can be aggregated.

A second option instead attempts intertheoretic comparisons of actions' choiceworthiness. Different techniques require progressively weaker comparability assumptions: one might attempt *cardinal* comparisons of the choiceworthiness of options across theories, *ordinal* comparisons that rank-order the choiceworthiness options across theories ("best option," "second-best option," etc.) , or else apply *bargaining* frameworks.

Several normative positions provide *cardinal* assessments of the desirability of options: for example, a total utilitarian might conclude that seeding independent colonies has far greater expected value compared to refraining from seeding independent colonies, and these two options can be cardinally assessed (the magnitude of the "far greater expected value" can be quantified). Intertheoretic cardinal assessments between positions with incommensurable scales might be made by normalising each theory's choiceworthiness assessments by their variance. By doing so, it may be possible to compare the degree to

which decisions are choiceworthy across several positions, such as the sentient positions (1) to (4), and thereby maximise expected choiceworthiness.

An objection to this process is that overall assessments of desirability across theories might be dominated by whichever theory claims the largest stakes - here, negative utilitarianism might hold that space colonisation carries enormous expected disvalue, and such vast magnitudes might dominate the intertheoretic comparison (Baker, 2024). One response questions whether this magnitude-dominance is necessarily a defective feature of the comparison (Wilkinson, 2021), while another argues that realistic background uncertainty about independent sources of value tends to wash out fanatical implications of expected-value reasoning in real-world decisions (Tarsney, 2025).

A more serious objection to cardinal rankings across theories is that several positions are best interpreted as providing ordinal, rather than cardinal rankings of options' choiceworthiness. The deontological position might rank restraint as being more choiceworthy than establishing independent colonies, even if it does not precisely quantify how much more choiceworthy the former is (MacAskill, Bykvist, and Ord, 2020, 153).[8] These theories might provide *ordinal* rather than cardinal comparisons, such that a voting-style aggregation rule such as Borda Scores may be appropriate in ordering options across theories. Each position is normalised by assigning credences and then multiplying that credence by the theory's Borda Score (the number of options that are less choiceworthy than the action, minus the options that are more choiceworthy, per MacAskill, Bykvist, and Ord, 2020, 73).[9] By attempting to maximise credence-weighted Borda Scores, decisions can be ordinally ranked by overall expected choiceworthiness. It should here be noted that ordinally ranking options provides far less information compared to cardinal rankings (it does not describe how much better a first-place option is relative to a second-place option), and Borda Scores are highly sensitive to the number of options considered during the assessment.

One might hold that neither cardinal nor ordinal comparisons can be made across normative positions, and choose to instead deploy parliamentary bargaining models to weight each theory. Analogous to parliaments, each theory supplies delegates to bargain over a debate, and the higher one's credence in a theory, the greater

[8] One potential problem is that deontological constraints might be better interpreted as categorical prohibitions, and that assessments of permissibility might not be commensurable with assessments of choiceworthiness. Ordinal ranks may offer shared intertheoretic grounds, and intertheoretic comparability might have to be assumed for reasoning under moral uncertainty to succeed.

[9] To attempt to maximise expected choiceworthiness, we would first need to assign credences to each theory. As the ten positions are not exhaustive, credences are conditional on the ideal assumption that the true theory is one of the ten.

the number of delegates (Greaves and Cotton-Barratt, 2023, 133). If one had a credence of 0.8 in negative utilitarianism (such that they were 80% confident that negative utilitarianism was correct), and only a 0.01 credence in nonhumanist perfectionism, then the former position would be assigned 80 times the number of delegates. The appeal of parliamentary models is that bargaining does not require well-defined units for intertheoretic comparisons (Greaves and Cotton-Barratt, 2023, 164). Further, low-credence theories might retain significance in the overall assessment by trading away their influence on issues they consider less important. Even at a modest credence, a position such as negative utilitarianism might influence one decision that it treats as catastrophically important (such as opposing unrestricted panspermia) by conceding influence elsewhere.

Parliamentary models may be helpful in inferring the critical decisions favoured by the whole set of normative positions when weighted by credence and adjudicated in rounds of bargaining. It is unclear, however, that bargaining would select for the "right" answer - the form of bargaining is unspecified (Greaves and Cotton-Barratt apply Nash Bargaining Solutions, yet other models might be preferred); for single-shot decisions such as "whether to seed planets with microbes," majority-credence positions might simply dominate (Newberry and Ord, 2021, 13); and just as when parties in actual parliaments threaten extreme positions to extract concessions, bargaining solutions might fatally subvert the preferences for each party to the bargaining process.

To conclude this discussion, whether it is feasible to produce an aggregate assessment of the desirability of space colonisation activities across disparate normative theories is itself a highly contested question. Arguably, it is important to attempt to formulate an aggregate assessment under conditions of high moral uncertainty - i.e. if the reader has non-negligible credences in multiple normative positions. Three techniques to reason under moral uncertainty require progressively weaker comparability assumptions: normalised cardinal ranking comparisons may be appropriate for certain consequentialist positions; ordinal ranking comparisons might extend comparability non-consequentialist positions while lowering the information supplied by the overall assessment; parliamentary approaches require neither cardinal nor ordinal comparability, yet solutions to bargaining processes are under-specified.

Finally, if we instead suppose that aggregate inter-theoretic assessments are infeasible or inappropriate, comparative axiological assessments remain informative. Conditional conclusions may still be described, and shared empirical cruxes underlying diverse desirability assessments may be identified. The desirability of

space colonisation remains highly contested depending on assumed normative positions, yet resolving empirical uncertainties - such as whether robust governance mechanisms might prevent risks of astronomical suffering, protect created biospheres, and preserve just conditions on Earth - goes some way in defusing disagreements. It is therefore crucial to investigate whether colonisation activities would be well-governed, consented-to, and reversible.[10]

[10] Such attributes are explored in (Peppiatt, 2026, 45-6).