Prudential Longtermism*

Johan E. Gustafsson and Petra Kosonen

ABSTRACT. According to Longtermism, our acts' expected influence on the expected value of the world is mainly determined by their effects in the far future. There is, given total utilitarianism, a straightforward argument for Longtermism due to the enormous number of people that might exist in the future, but this argument does not work on person-affecting views. In this paper, we will argue that these views might also lead to Longtermism if Prudential Longtermism is true. Prudential Longtermism holds for a person if and only if our acts' overall influence on that person's expected well-being is mainly determined by the acts' effects in the far future. We argue that (due to a small chance of anti-ageing and uploading) there could be an enormous amount of prudential value for some contemporary person in the far future and that value may be so large that it dominates their overall expectation of lifetime well-being.

According to

> *Longtermism*   Our acts' expected influence on the value of the world is mainly determined by their effects in the far future.[1]

Longtermism is counter-intuitive. It implies that our influence on the short-term, which we normally focus on, is outstripped by our influence on the far future. Yet, given a total utilitarian view of expectations, there is a straightforward case for Longtermism. Even a small possibility of a very large population living in the far future outweighs the importance of our acts' effects in the near future. Consequently, when evaluating acts, we can often simply ignore their short-term effects and focus on their effects in the far future.

It's less clear whether there is a similar case for Longtermism if we accept a person-affecting view, on which an outcome cannot be better

---

[1] The term 'longtermism' was coined by MacAskill and Ord (2020: 46, 306n27). See Greaves and MacAskill forthcoming for a defence of (strong) longtermism.

than some other outcome unless it is better for someone.[2] Our acts today affect not only the number and quality of future lives but also who will exist in the future — so that each act we can perform results in different people existing in the future due to the ripple effects of these acts.[3] So, if it can't be better or worse for someone to exist than not to exist, it seems that the only people we can make better off are those who already exist (and maybe people who will exist very soon).[4] In that case, if it's certain (or almost certain) that no one alive today will be alive in the far future, then person-affecting views lead to the rejection of Longtermism.[5]

But, in fact, there is a different path to Longtermism that is perfectly compatible with person-affecting views. Instead of total utilitarianism, this path appeals to

> *Prudential Longtermism*   Prudential Longtermism holds for person S if and only if our acts' overall influence on the expected prudential value for S is mainly determined by the effects of these acts in the far future.

If Prudential Longtermism is false for all currently existing people, then normative views on which only these people matter lead to the rejection of Longtermism.[6] Or, at least, they do so if we assume (as seems plausible) that a current person's well-being can only be determined by effects in the far future if they affect the well-being of some individual in the far future for whom the current person is justified in having prudential concern.[7]

---

[2] Narveson 1973: 80, Parfit 1984: 394–400, and Temkin 1987: 166–7.

[3] Parfit 1984: 351–5.

[4] It's controversial whether it can be better for a person to exist than not to exist. Williams (1973: 87), Parfit (1984: 487), and Broome (1993: 77) argue that it cannot, while Arrhenius and Rabinowicz (2015: 427–32) argue that it can.

[5] Bostrom 2003: 312. In contrast, asymmetrical views on which creating happy lives does not make the world better but creating unhappy lives does make the world worse, may lead to Longtermism. See Thomas forthcoming: 25–6. (See also Mogensen forthcoming and Steele forthcoming.)

[6] Many of the new insights from the recent flurry of research on effective altruism have yet to be applied to prudential concerns — an endeavour which we may call *effective prudentialism*. If we're effective when it comes to spending 10% of our income on altruistic causes, why be careless with the remaining 90%?

[7] Scheffler (2013: 73; 2018: 44) claims that, for many people, the value of their current activities depends on there being future generations continuing these activities — even though they accept that they, themselves, will die young. If so, Prudential Longtermism might be true for them, since their current well-being depends on the existence of other

In this chapter, we will explore whether Prudential Longtermism is true. Prudential Longtermism depends mainly on the feasibility of different forms of life extension. But, as we shall see, it also depends on what relation matters in survival and on how we should aggregate personal value in cases of fission — that is, cases in which there are multiple individuals in the future who are all related to a person (as the person is now) in the way that matters for survival.

We may distinguish between different strengths of Prudential Longtermism:

> *Strong Prudential Longtermism*   Strong Prudential Longtermism holds for person S if and only if our acts' overall influence on the expected prudential value for S is *overwhelmingly* determined by their effects in the far future.[8]

> *Weak Prudential Longtermism*   Weak Prudential Longtermism holds for person S if and only if our acts' overall influence on the expected prudential value for S is *mostly* determined by their effects in the far future.

If Weak Prudential Longtermism holds for someone, then the far future matters more in expectation than the near future for their prudential value. In contrast, if Strong Prudential Longtermism holds for someone, then the far future matters overwhelmingly more than the near future for their prudential value, and, for their prudential concerns, we can often simply ignore our acts' short-term effects and focus on the long-term effects.

In this chapter, we will discuss whether Weak and Strong Prudential Longtermism hold for some currently existing people and whether this means that even person-affecting views lead to (impersonal) Longtermism. It's clear that there are things we could do such that we would have no hope of any prudential value after the short-term. So, in our discussion, we will look for acts and technologies that may provide a lot of prudential value in the long-term. By performing such acts rather than the

---

people in the far future. Scheffler (2018: 53–7), however, denies that the existence of future generations after our deaths would provide us with prudential reasons — see also Greaves 2019: 138–140 for some objections to Scheffler on this point. In this chapter, we will not explore this posthumous route to Prudential Longtermism.

[8] MacAskill (2019) defines Strong Longtermism as the view on which we should be most concerned about the long-run outcomes, while Very Strong Longtermism is defined as the view on which the long-term outcomes are of overwhelming importance.

acts that offer no expectation of long-term prudential value for some person, our acts have an enormous influence on their expected prudential value. And then Prudential Longtermism holds for that person.

As mentioned, the case for Prudential Longtermism relies on the feasibility of extreme life extension.[9] There are a number of ways in which we might be able to extend our healthy lifespans. While these forms of life extension may be far-fetched, we will argue that, for some of them, even a small chance of them working is sufficient to support Prudential Longtermism. Hence, while we defend Prudential Longtermism, we are not claiming that any of these forms of life extension are likely to work.

1. Anti-ageing

Anti-ageing is the attempt to stop, or even reverse, ageing.[10] Research on anti-ageing has recently made some significant progress and may soon move on from non-human to human trials.[11] Could anti-ageing, by itself, lead to Prudential Longtermism? If it succeeds in stopping or reversing ageing, it could, of course, significantly lengthen our lives. But, even if we stop ageing, we may still die from other causes. Given a 0.13% chance of death per year (the proportion of people aged 30–31 who died in the U.S. in 2019), one has a 27% chance of surviving for 1,000 years and just a 0.00022% chance of surviving for 10,000 years. And one's life expectancy is $1/0.0013 \approx 770$ years.[12] This estimate assumes that the annual background risk of death (from injury or illness) won't change, and it doesn't take into account rare events, such as wars, global catastrophes, or existential risks.[13] Is 770 years a sufficiently long life expectancy to lead to

[9] Bostrom 2003: 312.

[10] See de Grey and Rae 2007 for an overview and defence of the feasibility and prudential value of anti-ageing, and see Bostrom 2005 and Bostrom and Ord 2006: 676–7 for further defences of the desirability of lengthening our healthspans.

[11] For an optimistic overview of recent advances in anti-ageing research, see Partridge et al. 2020. Crimmins (2015: 908–9) offers a less optimistic take, claiming that the necessary interventions may need to be done at a very young age. So, even if anti-ageing will be invented in our lifetimes, it may be too late for current adults. In other words: if you can read this, it may be too late for you.

[12] Based on data from Arias and Xu 2022: 10. Similarly, Bostrom and Roache (2007: 124) estimate that, if we lived at the mortality rate of someone in their late teens or early twenties, our life expectancy would be around 1000 years.

[13] Ord (2020: 167) estimates that the risk of human extinction within the next 100 years is already 1/6.

Prudential Longtermism?

Let the next 100 years constitute the short-term, and let the long-term start thereafter.[14] And let us assume (somewhat arbitrarily) that a technology leads to Strong Prudential Longtermism if and only if, due to this technology, a person's expected number of life years in the long-term (the period starting after the next 100 years) is at least 10,000 times as great as their expected number of life years in the short-term (the next 100 years). This will be true if their life expectancy is at least 1 million (plus 100) years, assuming that the person will certainly live for 100 years.[15] If there were such a technology, then it would be plausible that our acts' overall influence on the expected well-being of some currently existing person is overwhelmingly determined by our acts' effects in the far future. Next, let us assume that a technology leads to Weak Prudential Longtermism if and only if a person's expected number of life years in the long-term (the period starting after the next 100 years) is greater than that person's expected number of life years in the short-term (the next 100 years). Then, assuming that the long-term does not provide opportunities for far greater or far lower welfare per unit of time than the short-term, it's likely that our acts' overall influence on some currently existing person's expected well-being is mostly determined by their effects in the far future.

How high must our credence in anti-ageing working be in order for it to lead to Weak Prudential Longtermism? By anti-ageing working for a person, we mean anti-ageing being successfully used by them. With p being the constant probability of death each year if anti-ageing works (which we have assumed to be 0.13%), we find that the expected years of life in the short-term (that is, the next 100 years) if anti-ageing works is

$$\sum_{n=1}^{100} (1 - p)^n \approx 93.7.$$

[14] In Greaves and MacAskill (forthcoming), the far future is 'everything from some time t onwards, where t is a surprisingly long time from the point of decision (say, 100 years).'

[15] Temkin (2008: 202–4) argues that an extremely long life may get boring after a while, noting that he has listened to his favourite music (mostly late 60s and early 70s rock) so much that it no longer gives him much pleasure. (For a similar complaint, see Williams 1973: 90.) Note that, if our lives would inevitably become more or less neutral in the far future, then we would not make a difference to how much well-being we would have in the far future even if we develop anti-ageing technology. So might we, like Temkin, eventually run out of new pleasures? That seems unlikely. There is a simple solution to Temkin's predicament: Try some new music.

Let q be the probability of anti-ageing working. And assume that a person's current life expectancy without any new life-extension technology is 50 years (the U.S. life expectancy at age 30 in 2019).[16] Now, anti-ageing alone leads to Weak Prudential Longtermism if

$$\left(\frac{1}{p} - 93.7\right)q > 93.7q + 50(1 - q).$$

Hence anti-ageing leads to Weak Prudential Longtermism if q, the probability of anti-ageing working, is greater than 8%. Then, the person's expected number of life years in the long-term is greater than their expected number of life years in the short-term.[17]

But anti-ageing alone does not lead to Strong Prudential Longtermism (as we have defined it). Even assuming that anti-ageing is guaranteed to work, the expected number of life years in the long-term is less than 8 times greater than the expected number of life years in the short-term, given a 0.13% yearly chance of death. Of course, we may be able to decrease our yearly risk of death and thereby improve our chances of survival significantly. In order to get 10,000 times as great expectation of number of life years in the long-term as in the short-term, we need the annual risk of death to be at most one-in-a-million. But this, of course, assumes that anti-ageing works. Since there is uncertainty about the feasibility of anti-ageing, the annual risk of death needs to be even lower for anti-ageing alone to lead to Strong Prudential Longtermism.

2. Cryonics

Cryonics is the process of storing a person's brain (or whole body) at a very low temperature after their legal death in the hope that they will one day be revived. A current way to cryopreserve a brain is through vitrification, which hardens water like glass without crystal formation that would cause damage to cells. The brain is then kept cool with liquid nitrogen. The hope is that this process will preserve the brain without further tissue degradation and that medical science will eventually make advances that allow the stored brains to be revived (and repaired) back to a healthy

---

[16] Using data from Arias and Xu 2022: 3.

[17] But anti-ageing alone need not give us Weak Prudential Longtermism if we accept the Multiplicative View of Continuity Strength (discussed in section 3) and the weight of Relation R holding between consecutive person-slices is less than one.

life.[18]

One worry about cryonics is whether it can preserve memories.[19] Many philosophers believe that psychological continuity is what matters in survival.[20] On this view, an outcome is as bad as death for a person unless they are psychologically continuous with someone in that outcome. Psychological continuity in turn consists in overlapping sequences of psychological connections. And these connections are usually taken to be memory relations, that is, the relation of the current person's experiences being remembered by the future person.[21] So, on these views, cryonics does not preserve what matters in survival if it fails to preserve one's memories.

Yet there are other candidates for what matters in survival. Some believe that *physical* continuity is what matters.[22] On these views, an outcome is as bad as death for a person unless that person has the same brain (or enough of the same brain) as someone in that outcome. Thus, on these views, cryonics could preserve what matters in survival even if it fails to preserve memories (or any other psychological connections) — as long as it's possible to revive the same spatio-temporally continuous brain.

Does cryonics in combination with the technology to revive a cryopreserved brain lead to Strong Prudential Longtermism? Even if cryonics combined with such a technology leads to a successful revival, it is still open to worries about fatal injuries that permanently destroy the brain after the revival. So, even if it's possible to revive the spatio-temporally continuous brain after cryopreservation and this brain could be given a new body, that brain may still be damaged beyond the possibility of revival. The annual risk of brain destruction (during those years in which the brain is not cryopreserved) would have to be at most one-in-a-million in order to get at least one million life years in expectation. This bar for

---

[18] Merkle 1992: 6; 1994: 16. The feasibility of cryonics is controversial. While some scientists think that it could work (see, for example, Benford 2005), others claim that it's impossible (see, for example, Miller 2004). For a defence of the practice of cryonics, see Shaw 2009.

[19] Doyle 2018: 124. Vita-More and Barranco (2015: 458), however, claim to have made progress in preserving long-term memory in worms after cryopreservation.

[20] According to Minerva 2018: 10, the dominant view among supporters of cryonics is that a person is fundamentally the information stored in a brain.

[21] Parfit (1984: 205), however, suggests that other psychological relations may also matter. Plausibly though, if memory connectedness does not hold, this is likely to be accompanied by the rupture of other connections.

[22] Nagel 1986: 40 and Unger 1990: 109.

the risk of brain destruction is still too low for Strong Prudential Longtermism to be true for anyone. Hence cryonics in combination with anti-ageing and the technology to revive a cryopreserved brain at most gives us Weak Prudential Longtermism. Still, cryonics (like anti-ageing) might buy us time for finding better ways of extending life.

3. Uploading

Uploading (also known as whole-brain emulation) is the process of scanning a person's brain and loading the information onto a computer, where the brain is then simulated.[23]

A standard worry about uploading is whether the simulation will be conscious.[24] A zombie simulation would not (assuming hedonism) have any well-being, so it would be prudentially worthless, intrinsically. Another worry is whether the upload would be identical to the current person, that is, whether the current person would be identical to their simulated self.[25] A more pressing worry, however, is whether the current person would stand in the relation that matters in survival to their simulation. The views of personal identity on which one could plausibly be identical to one's simulation are reductionist views where personal identity just consists in an impersonal mental relation holding uniquely.[26] Here, an impersonal relation is a relation that can be completely described without mentioning people. But, if personal identity can be reduced to an impersonal relation (such as psychological continuity) holding uniquely, it seems that we should also care about this relation when it holds from one to many rather than only in the case it holds from one to one.[27]

The most influential reductionist view is that psychological continuity is what matters in survival. As mentioned, psychological continuity is the holding of overlapping sequences of psychological connectedness. Psychological connectedness is a direct psychological connection between a person at one time and a person at another time, such as the person at the latter time remembering (or quasi-remembering) the

---

[23] See Sandberg and Bostrom 2008: 7-15, for an overview of uploading.

[24] Chalmers 2010: 44–8.

[25] Aaronson 2016: 210–1 and Chalmers 2010: 48–63.

[26] Parfit 1984: 263. Actually, the structure of the account must be somewhat more complicated. See Gustafsson 2019: 2314–5.

[27] Parfit 1971: 4–14; 1984: 256, 261–4 and Gustafsson 2018: 745–50.

experiences of the person at the earlier time.[28]

To discuss psychological continuity and the relation of what matters in survival, we will adopt a perdurance framework in which we analyse persistence in terms of person-slices, that is, instantaneous temporal parts of people.[29] (But our discussion could also be done in an endurance framework, changing what needs to be changed.) We will represent psychological connectedness by Relation C, defined as follows:

> Person-slice x is C-related to person-slice y (xCy) $=_{df}$ x is psychologically connected to y with the right kind of cause and x is either simultaneous with y or earlier than y.

We will represent psychological continuity by Relation R, which we define in terms of the holding of overlapping chains of C-relations:[30]

> Person-slice x is R-related to person-slice y $=_{df}$ either xCy or yCx, or there are person-slices $z_1, z_2, \ldots, z_n$ such that either
>
> (i) $xCz_1, z_1Cz_2, \ldots, z_{n-1}Cz_n, z_nCy$ or
>
> (ii) $yCz_1, z_1Cz_2, \ldots, z_{n-1}Cz_n, z_nCx$.

Both Derek Parfit and David Lewis wobble a bit on whether to put the emphasis on continuity or on connectedness — that is, whether it's Relation C or Relation R that matters (or both).[31] The distinction is crucial for determining the amount of prudential value someone gets from uploading.

A VERY LONG SIMULATION

One reason to think that uploading may lead to Prudential Longtermism is that the uploads can live on for a very long time.[32] If the simulations of a

---

[28] Quasi-remembering is just like remembering except that the remembered person needn't be the same as the remembering person. See Shoemaker 1970: 271.

[29] Brink 1992: 215–6. To avoid overlap between stages, we rely on person-slices rather than person-stages, which need not be instantaneous. For person-stages, see Perry 1972: 467 and Lewis 1986: 202. We may wish to allow that person-slices, rather than being instantaneous, have the minimal duration necessary to be able to have well-being. But, if so, we need to individuate the slices in a way that avoids overlap between slices.

[30] McMahan 2002: 50.

[31] Parfit 1971: 21; 1984: 262 and Lewis 1976: 18.

[32] Dyson (1979: 456) suggests that a finite amount of physical energy could be used to simulate an infinite amount of subjective time.

currently existing person *S* gradually change their psychology over time, they may eventually stop being *C*-related to *S* as *S* is now, even though they would still be *R*-related to *S*.[33] Since prudential concern is plausibly forward (rather than backward) looking, the simulations need not have any special interest in continuing to be directly psychologically connected to *S*.[34] So we may suspect that they will gradually let go of their memories of *S* in order to make room (in computer memory) for more useful knowledge.[35] Hence, if Relation *C* is what matters, it seems that uploading does not lead to Prudential Longtermism in virtue of a very long-lasting simulation. But, if Relation *R* is what matters in survival, it seems that, as long as the simulation is kept running, one's relation to one's simulation contains what matters. And, if civilization survives and people have some interest in keeping the simulation running, then the simulation may run for a very long time.

Assuming that *S*, as *S* is now, is *R*-related to a large number of person-slices of a long-lasting simulation, how much prudential value does this provide for *S*? This depends on three factors: (i) how much *S*'s relation to each of these person-slices matters, (ii) how well-off these person-slices are, and (iii) how the well-being of these person-slices should be aggregated.

Let a *life-path* be a maximal aggregate of person-slices that are related by what matters to each other, that is, an aggregate of person-slices such that (i) each slice in the aggregate is related by what matters to all slices in the aggregate and (ii) no person-slice that is not in the aggregate is related by what matters to all slices in the aggregate. The important thing about life-paths is that they are unified in the sense that the relation that matters does not branch, as all person-slices in a life-path are related by what matters to all others in that life-path. (On some views of personal identity, a life-path is a person.[36] But we do not need to assume this.)

[33] Lewis 1976: 29–31.

[34] Parfit (1984: 174–7), however, challenges this bias towards the future.

[35] But, if the simulations accept evidential decision theory, they may wish to keep memories of their earlier person-slices, because letting go of those memories would be evidence that the later person-slices will also choose to let go of their memories. (For evidential decision theory, see Jeffrey 1965: 1–6; 1983: 1–6, Gibbard and Harper 1978: 157, and Ahmed 2014: 43–6.) Alternatively, could one try to cultivate a false belief in backward-looking prudential concern? It seems that, if one can tell that a philosophical view is implausible, then these descendant simulations would be able to do so too.

[36] Our definition of a life-path corresponds to Lewis's (1976: 22) definition of a continuant person.

Now, regarding the aggregation of the well-being of the future person-slices, consider

> *The Single Life-Path Total View*   Within a single life-path, the overall prudential value of a risk-free prospect for person S now is the sum total, for all future person-slices within that life-path, of the well-being of that slice multiplied by the weight of the R-relation between that slice and S (as S is now).

On this view, a person's future momentary well-being is added up, in proportion to the weights of the R-relations, to get the prudential value of their future. The Single Life-Path Total View implies

> *The Single Life-Path Repugnant Conclusion*   For any possible life-path in which each person-slice has very high well-being and is R-related to person S (as S is now), there is a possible life-path that is better for S even though each of its person-slices has barely positive well-being.

This conclusion implies that, for any number of years that a person could live at a high momentary well-being level, there is some number of years, during which they have barely positive momentary well-being, that is better for them.[37] The Single Life-Path Total View also entails the following variant where the weights of the R-relations are different but well-being is held constant:

> *The Weighted Single Life-Path Repugnant Conclusion*   For any possible life-path in which each person-slice has positive well-being and is strongly R-related to person S (as S is now), there is a possible life-path that is better for S even though each of its future person-slices is barely R-related to S (as S is now) and, in both cases, the person-slices within those life-paths have the same positive well-being.

While these conclusions may seem counter-intuitive, they seem less so than the corresponding Repugnant Conclusion in population ethics.[38]

---

[37] See McTaggart 1927: 452–3, Parfit 1986: 160, Crisp 1997: 24–5, and Temkin 2012: 119.
[38] Parfit 1984: 388.

Moreover, we can defend these single life-path conclusions with a mere-addition argument:[39] Adding a long life that is at each point minimally positive in well-being to a person's lifespan seems to be at least as good for them as their life without that addition. Then, making that person's life equal in quality throughout while increasing the average momentary level of well-being slightly seems to be better for them. Then, by the transitivity of *at least as good as*, we find that the end result — that is, a life that is at all times barely worth living — would be better for the person than their current lifespan (no matter how good their current lifespan is).[40]

But it's less obvious how to weigh the importance of being *R*-related to a person-slice. Relation *C* has a straightforward weighting: the proportion of how much of the earlier person-slice's psychological state the later person-slice shares or remembers. Since Relation *R* holds in virtue of overlapping sequences of *C*-related person-slices, it's compelling to adopt

---

[39] This argument is analogous to the Mere-Addition Paradox in Parfit 1984: 419–41.

[40] If one is tempted to resist the Single Life-Path Total View, one could adopt

> *The Single Life-Path Average View*   Within a single life-path, the overall prudential value of a risk-free prospect for person *S* now is the sum total, for all future person-slices within that life-path, of the well-being of that slice multiplied by the weight of the *R*-relation between *S* (as *S* is now) and that slice divided by the sum total of all the *R*-relations weights. (Note that all person-slices are assumed to last equally long.)

On this view, a person's future momentary well-being levels are averaged over (while taking into account the weights of the *R*-relations) to get the prudential value of their future. This view, however, implies

> *The Single Life-Path Masochistic Conclusion*   It can be better for person *S* if, within a single life-path, *S* (as *S* is now) were related by what matters in survival to a small number of additional person-slices with negative well-being than if *S* (as *S* is now) were related by what matters in survival to a large number of additional person-slices with positive well-being (other things being equal).

This conclusion follows, because the person's average momentary well-being might be decreased less by the addition of the person-slices with negative well-being than by the addition of the person-slices with positive well-being. Another problem for the Single Life-Path Average View is that, if it is future-oriented, the prudential value of immediate death is undefined as there would not be any future person-slices whose well-being can be averaged over. But, if we take the average over one's life-time instead, then we get an analogous problem to the Egyptology objection to average utilitarianism: what happened in someone's distant childhood matters for which future is best for them. See McMahan 1981: 115 and Parfit 1984: 420.

the following view:[41]

> *The Multiplicative View of Continuity Strength*   Let a
> *weight-product* of a sequence of C-related person-slices be equal
> to the product of the weights for each C-relation in the sequence.
> The weight of Relation R holding between person-slices x and y
> is equal to the maximum weight-product of any sequence xCy or
> yCx or a sequence via person-slices $z_1, z_2, \ldots, z_n$ such that either
>
> (i) $xCz_1, z_1Cz_2, \ldots, z_{n-1}Cz_n, z_nCy$ or
>
> (ii) $yCz_1, z_1Cz_2, \ldots, z_{n-1}Cz_n, z_nCx.$

Note that, between two person-slices, there might be lots of sequences of overlapping C-relations and that the sequence with the greatest weight-product need not be the longest — it may even be the sequence consisting of a single direct C-relation between the two person-slices.

Does this view lead to Strong Prudential Longtermism given a successful upload with a long-lasting simulation? The trouble is that, once we allow the C-relations between the simulated person-slices to have weights that are less than 100%, the sum of the weights of the R-relation for all person-slices will converge relatively quickly, assuming that current memories (or whatever the R-relation consists of) are not retained at a higher rate than future memories. Each person-slice has prudential reasons to prefer being remembered by the next person-slice, so they would not opt to be forgotten by their immediate successor. But it seems that person-slices need not have any prudential reason to prefer that their predecessors are remembered by the next person-slice. So it seems that person-slices may opt to forget earlier person-slices in order to free up resources for more important information (or additional simulations). Let us therefore assume, to make the calculation simple, that person-slices only remember their immediate successor person-slice. Let each person-slice of the simulation be a year long. And suppose that the well-being of each person-slice is constant at u. Let the weight of each C-relation be w. Then, given the Multiplicative View of Continuity

---

[41] This view entails McMahan's (2002: 50) view that prudential concern is transitive: If the relation that matters holds to some extent between person-slice x and person-slice y and to some extent between y and person-slice z, then it holds to some extent between x and z.

Strength, the prudential value of an x years long simulation is

$$\sum_{i=1}^{x} uw^i = \frac{uw(w^x - 1)}{w - 1}$$

As the simulation lasts longer, this converges to

$$\sum_{i=1}^{\infty} uw^i = -\frac{uw}{w - 1}$$

To see that this does not favour Strong Prudential Longtermism, note that (given a positive well-being u and given that the weight w for the C-relations is positive and not greater than 100%) the infinite number of years after the first 100 years do not contribute 10,000 times more to the prudential value of the future than the next 100 years unless 99.9999% of each person-slice's psychology is retained each year.[42]

Would it be in each person-slice's interest that the next person-slice of the simulation remembers them to this extreme extent? It may seem that it would, because the more the next person-slice remembers them, the more the next slice (and the future) matters to them. But, if each slice needs to remember the last one completely, it seems that the simulation would constantly need more memory in order to store new knowledge. (Computational resources could also be used to create more simulations.) So it would make sense at some point to forget the last person-slice to some extent. But, if so, a long-lasting simulation does not (by itself) lead to Strong Prudential Longtermism.

Another potential way in which a long-lasting simulation may lead to Strong Prudential Longtermism is if the well-being levels of the person-slices of the simulation gradually get better. Even if the sum of the weights for the R-relations converges, it might still be that the overall prudential value increases faster and faster. With the addition of technological ad-

---

[42] This results in a form of discounting of the future. But it is not a pure-time preference of the kind Sidgwick (1907: 381), Ramsey (1928: 543–4), Rawls (1971: 293; 1999: 259), and Parfit (1984: 125–6) object to. Yet Ahmed (2020) does object to this kind of psychological discounting. It's unclear, however, why we should accept his (2020: 247) Stationarity assumption that one takes at all times the same attitude towards well-being at the same distance in the future. If, on Monday, one knows that they will lose a lot of memories on Thursday and lose very few memories before then, then one plausibly cares more on Monday about one's Wednesday well-being than one will care on Wednesday about one's Friday well-being.

vances over time, we may be able to achieve increasingly higher welfare, and this might offset the decreasing weights of the R-relations to these distant person-slices.

BRANCHING SIMULATIONS

Earlier, we distinguished the view that some relation matters in survival from the view that personal identity matters in survival — even if personal identity only consists in the former relation holding uniquely (that is, without branching). When we assess the prudential value of uploading, the difference between these views matters a great deal. The reason it matters is that, once we have created a simulation of someone's brain, we can create many more.[43]

If we allow for branching in the relation that matters, we allow that someone can stand in the relation that matters to two (or more) simultaneous person-slices (which do not stand in the relation that matters to each other). But how should we aggregate the well-being of future person-slices in branching cases — that is, in cases of fission?[44]

Suppose that a person S will undergo uploading and that either (A) one simulation will be created and it will enjoy four years of high momentary well-being, or (B) that simulation and a separate simulation will be created and each of these simulations will enjoy three years of high momentary well-being at the same momentary well-being level as in A (Ω denotes non-existence):

	S_1	S_2
A	4	Ω
B	3	3

Consider next, expanding the additive approach of the Single Life-Path Total View to cases involving multiple life-paths,

---

[43] See Dainton 2012: 56 for a discussion of fission through multiple uploads.

[44] One benefit of fission is that it allows one to become multi-planetary in the sense that one could stand in the relation that matters in survival both to future people on Earth and simultaneous (or space-like separated) future people on Mars. This allows one to survive a catastrophe that eliminates all life on one of these planets. This is, of course, analogous to the quest to safeguard humanity as a whole by becoming multi-planetary. See Sagan 1994: 371, Parfit 2017: 436, and Ord 2020: 392–3n16, but compare Ord 2020: 194.

The Prudential Total View   The prudential value of a risk-free prospect for person *S* is equal to the sum total of the well-being of every person-slice that *S* (as *S* is now) is related to by the relation that matters, where the well-being of each slice is weighted by the strength of that relation.[45]

On this view, *S* would be better off if two three-year simulations were created instead of one four-year simulation, that is, *B* is prudentially better than *A*.

Let *a person's life-paths* be the life-paths that have that person's current person-slice as a member. The Prudential Total View entails the following conclusion:[46]

The Prudential Repugnant Conclusion   For any outcome in which each of person *S*'s life-paths has a great prudential value for *S*, there is an outcome that is better for *S* even though each of *S*'s life-paths in that outcome has a barely positive prudential value for *S* (and, in both outcomes, the person-slices within the life-paths have the same weights for the *R*-relations).

In the case of uploading, this conclusion implies that, for any number of simulations of *S* with very high well-being, there is a prudentially better outcome for *S* that contains a much larger number of simulations of *S*, each with a barely positive well-being level (while holding the weights of the *R*-relations fixed).

The Prudential Total View also entails the following variant, where the weights of the *R*-relations are different (but well-being contained in each life-path is held constant):[47]

The Weighted Prudential Repugnant Conclusion   For any outcome in which all the future person-slices of person *S*'s life-paths are strongly *R*-related to *S* (as *S* is now), there is an outcome that is better for *S* even though all the future person-slices of *S*'s life-paths in that outcome are barely *R*-related to *S* (as *S* is now) and the sum total of well-being of person-slices in each life-path is the same in both outcomes.

[45] Holtug (2001: 55; 2010: 118) presents a person-focused (rather than person-slice-focused) prudential total view. And Ross (2014) argues against a similar view.

[46] Gustafsson and Kosonen forthcoming. We assume here that well-being can be represented by a real-valued function.

[47] See Holtug 2001: 60.

In the case of uploading, this conclusion implies that, for any number of simulations that are at all times strongly R-related to S as S is now, there is a prudentially better outcome that contains a much larger number of simulations that are barely R-related to S as S is now (holding the well-being of the simulations fixed).

We can contrast the Prudential Total View with an average view. The latter is slightly more complicated than one might think, since we still would like to maintain a sum-total view concerning the aggregation of momentary well-being over time (within one life when there is no fission).[48] To do so, we will introduce some terminology. As before, let a life-path be a maximal aggregate of R-related person-slices — that is, an aggregate of person-slices such that (i) each slice in the aggregate is R-related to all slices in the aggregate and (ii) no person-slice that is not in the aggregate is R-related to all slices in the aggregate.[49] Let a *successor* to a person-slice x be the person-slice that is next after x in a life-path of which x is part. Let a *fission-slice* be a person-slice with multiple successors.

> *The Prudential Average View*   Evaluate the prudential value of each life-path by the Single Life-Path Total View. Assume that fission-slices are followed by a chance node with an equal probability of being followed by each of that slice's successors. Hence we transform prospects with fission into prospects of uncertainty. The prudential value of a prospect for person S is equal to S's expected well-being in the transformed prospect.[50]

On this view, we treat the prospect of the two three-year simulations as if it were a fifty-fifty lottery between each of the two simulations being implemented on its own without the other. Hence, on the Prudential Average View, the prudential value of the two three-year simulations is the

[48] See note 40 for an argument against the average view concerning the aggregation of momentary well-being over time.

[49] Lewis 1976: 22.

[50] Tappenden 2011: 302. Another way to formulate the Prudential Average View would be to average over the well-being of all life-paths. This, however, would imply that, if person S first splits into two and much later one of the fission products splits multiple times (while the other does not), then that fission product's well-being (even before the later splits) would have overwhelmingly more influence on S's prudential value, because it is part of multiple life-paths. Thus this results in a form of double counting of well-being.

same as the prospect of a single three-year simulation — which is worse than the single four-year simulation.

Which of these answers is more plausible? Combining Parfit's Division Argument and his Mere-Addition Paradox, there is a straightforward argument for the answer of the Prudential Total View.[51] Consider, in addition to A and B, a third prospect A^+ that is just like A except that a second simulation is also implemented and this additional simulation has the same momentary well-being level as the first simulation but is only run for one year:

	S_1	S_2
A	4	Ω
A^+	4	1
B	3	3

It seems that, if simulation S_1 in A provides what matters in survival, then the same simulation in A^+ should also provide what matters in survival. The only difference in A^+ is that, in addition to S_1, there is another simulation to which S also stands in the relation that matters. So, in terms of what matters in survival, A^+ should be at least as great a success as A.[52] Consequently, A^+ should be at least as good as A for S. (Hence we should reject the Prudential Average View.[53]) Next, compare A^+ and B. Prospect B differs from A^+ in that S_1 lives for one year less but S_2 lives for two more years. Given that S stands in the relation that matters to *both* simulations, in terms of prudential value, the two extra years for S_2 in B should outweigh the single extra year for S_1 in A^+. So B is better than A^+ for S. Then, by the transitivity of *at least as good as*, we find that B is better than A for S.[54] Changing what needs to be changed, this argument also shows that we should accept the Prudential Repugnant Conclusion and the Weighted Prudential Repugnant Conclusion.[55]

---

[51] Parfit 1984: 419–26.

[52] Parfit 1971: 5; 1984: 256, 261–2; 1993: 24–5; 1995: 42.

[53] This argument is adapted from Gustafsson and Kosonen forthcoming.

[54] The transitivity of *at least as good as* can be taken to be an analytic principle of logic. See Broome 2004: 50–63. Or it can be defended with a money-pump argument. See Gustafsson 2022a: 39–44.

[55] The Prudential Average View also, implausibly, entails

> *The Masochistic Conclusion*   It can be better for person S if some number of additional life-paths with negative prudential value are created instead of some number of life-paths with positive prudential value being created (other things

Given that we adopt the Prudential Total View, rather than the Prudential Average View, we seem to have a route to Strong Prudential Longtermism. If we create not just one simulation of some currently existing person S but a large number of simulations, S's prudential value from these simulations increases in proportion to the number of simulations. Moreover, each one of these simulations is in much the same situation, as they also increase their prudential value from the future the more simulations there will be of them. And, in turn, these further simulations are in much the same situation, as they can increase their prudential value by creating even more descendant simulations. Hence it seems that we would get an explosion of more and more simulations that are R-related to S as S is now.[56] Since this increase in the number

---

being equal).

(This is a one-person counterpart to the Sadistic Conclusion. See Arrhenius 2000: 54. See also note 40 for a life-path version.) To see how the Masochistic Conclusion follows, consider prospects A and B. In prospect A, there are three separate simulations of person S: S_1 has a well-being of 13, whereas S_2 and S_3 have a well-being of 1. In prospect B, there are just two simulations of S: S_1 has a well-being of 13 (just like in A) and S_2 has a well-being of -1:

	S_1	S_2	S_3
A	13	1	1
B	13	-1	Ω

Here, the Prudential Average View entails that, for S, the prudential value of A is $(13 + 1 + 1)/3 = 5$ and the prudential value of B is $(13 + (-1))/2 = 6$. Thus it entails that B is better than A for S — which is an instance of the Masochistic Conclusion.

[56] Furthermore, note that, once a scan has been made of a person S, any replicas created from that scan no longer stand in the relation that matters in survival to S as S is after that scan. Or, at least, the replicas from the old scan wouldn't do so if, as seems plausible, the relation that matters in survival is temporally ordered (like Relation R is defined in this chapter). Some people take the relation that matters to be temporally unordered. For example, we could define a temporally unordered variant of psychological continuity as follows:

Person-slice x is C'-related to person-slice y $(xC'y) =_{df} x$ is strongly psychologically connected to y with the right kind of cause.

Person-slice x is R'-related to person-slice $y =_{df} xC'y$ or there are person-slices $z_1, z_2, \ldots, z_n$ such that $xC'z_1, z_1C'z_2, \ldots, z_{n-1}C'z_n, z_nC'y$.

The trouble is that, in a standard fission case where Wholly splits into Lefty and Righty, it seems plausible that Wholly is C'-related to each of Lefty and Righty. But it does not seem plausible that Lefty has what matters in survival to Righty, even though Lefty is R'-related to Righty. (See Gustafsson 2021: 509.) Given that the relation that matters is temporally ordered — that is, like Relation R rather than Relation R' — each person-

of simulations will outweigh the diminishing weight of the R-relation between S, as S is now, and the simulations as they get more distant from S, S will (at least in expectation) get most of their prudential value from this enormous number of simulations in the far future. Hence, if we have a sufficiently high credence in uploading working and S having sufficiently many future branches (and in us being able to non-negligibly affect the welfare levels of those branches), then we get Strong Prudential Longtermism.

Given this explosion in the number of simulations, there will be a similar explosion in the demand for computational resources. This would put everyone in competition with everyone else for any available computational resources. Could this competition be avoided? It seems that it could. If the relation that matters in survival can split into multiple branches, it seems that it could also merge from many branches into one.[57] In the case of Relation R, this would happen when a person-slice is psychologically connected (that is, remembers) earlier person-slices from multiple branches.[58] Then each simulation that will merge is R-related to all descendants of the merged simulation. This solution is structurally the same as Parfit's example of beings who merge and divide every autumn and spring:[59]
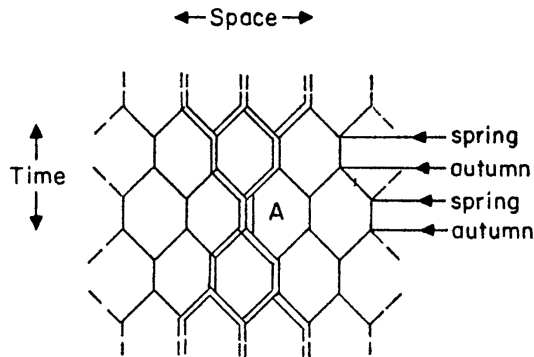
With these regular intervals of merging and splitting, everyone's pruden-

---

slice has an incitement (given a prudential motivation) to get another scan done and create even more replicas (which fuels the explosion of replicas further).

[57] Yet it may be harder to merge than to split. See Hanson 2016: 51.

[58] Bostrom (2014: 61) observes that digital minds might want to share memories to increase their knowledge faster. They may be able to save computational resources by storing just one instance of the memories, even though they can all access those memories. This raises the question of where a mind ends. See Clark and Chalmers 1998.

[59] Parfit 1971: 22; 1984: 303.

tial interests would overlap to a very large extent with those of everyone else.

So far in our discussion of uploading, we have assumed that psychology is what matters in survival. If (i) psychological continuity is what matters, (ii) uploading is feasible, (iii) uploading preserves psychological continuity, and (iv) simulations of brains would be conscious, it follows that one would have what matters in survival to a simulation from one's uploaded brain. Or, at least, it follows if this continuity need not have its normal cause: being caused by the continued existence of one's brain.[60]

But there is at least one rival to the psychological view that may also allow that uploading provides what matters. On the phenomenal view of what matters in survival, someone has what matters in relation to a future person-slice if and only if they are phenomenally continuous with that future person-slice.[61] Phenomenal continuity is, basically, the relation of partaking of the same stream of consciousness. In the same way as psychological continuity is the holding of overlapping sequences of psychological connectedness, phenomenal continuity is the holding of overlapping sequences of phenomenal connectedness. Phenomenal connectedness, in turn, is the relation of experienced togetherness — that is, the relation of one experience being experienced together with another experience in the same conscious state. This relation can hold between experiences at a time, such as your current visual experience and your current auditory experience. But this relation of experienced togetherness can also hold over time. For example, when you are listening to music and you hear one note transition into the next, or more generally, each one of your experiences flows into the next.[62] On this phenomenal view, a simulation provides what matters in survival for a person S if and only if a stream of consciousness S currently partakes of will include the experiences of the simulation.

If simulations can be conscious and have experiences at all, could we get a human person's stream of consciousness to transfer to a simulation? Barry Dainton suggests that we can. By gradually replacing more and more of the person's brain with functionally equivalent digital silicon-

---

[60] Parfit (1984: 283–7) defends this kind of view where there are no restrictions on how the psychological continuity is caused.

[61] Gustafsson 2021: 513n28.

[62] Dainton and Bayne 2005: 553–4. Phenomenal continuity can also hold between experiences that are separated by a period of unconsciousness, such as dreamless sleep. See Gustafsson 2011: 291–4.

based parts, he suggests that their stream of consciousness would continue intact.[63] Then, if done gradually, uploading may still provide what matters in survival even on the phenomenal view.

4. Biological uploading

One worry about uploading is that, as mentioned earlier, computer simulations might not, for all we know, be conscious. Yet, if we have a detailed scan of someone's brain, we might be able to create a new human complete with their psychology through what we will call *biological uploading*. The standard implementation of this idea is the teletransporter: A person steps into a machine on Earth which scans their body and then eliminates it and sends the scanned information to Mars, where a biological copy of the person is generated from the scan.[64]

If the creation of biological replicas with someone's psychology is possible, it may lead to Prudential Longtermism in much the same way as uploading. Or, at least, it may do so on the psychological view of what matters in survival. In contrast, on the phenomenal view, it seems less likely that the scanned person would have what matters in relation to their replicas, because it's implausible that the replicas' experiences would be part of the same stream of consciousness as the scanned person's experiences.

But we may be able to use a similar gradual approach to get it to work. It seems that, if a person S's brain is split in two, S would have phenomenal continuity to both halves. These brain halves can then be placed in two separate bodies and complemented with a replica of the other half. The result should be two people who each have a complete brain and who are both phenomenally continuous with S. Then we repeat this, if necessary, to generate the desired number of replicas.

(Actually, if we don't care about keeping S's psychology intact, this last procedure does not, fundamentally, need uploading. If we don't care about psychology, the brain halves needn't be combined with replicas of the other half — any compatible brain half will do.)

---

[63] Or rather, on Dainton's (2012: 55) view, the person's *capacity* for a continuous stream of consciousness would continue intact. See also Chalmers 2010: 52–5 for a discussion of gradual uploading.

[64] Parfit 1984: 199.

5. Prudential and Empirical Uncertainty

So far, we have seen that uploading and biological uploading can lead to Strong Prudential Longtermism. This requires that either psychological or phenomenal continuity is what matters in survival and, moreover, that we aggregate well-being in the way prescribed by the Prudential Total View (or some similar additive principle). These assumptions are plausible. The prudential analogue of the Mere-Addition Paradox is compelling and suggests the Prudential Total View or something similarly additive. That the relation that matters in survival is some mental relation (psychological or phenomenal) is also compelling. The reason why our brains and bodies seem so important in survival is that they are needed (so far) for mental continuity — but they are not what fundamentally matters. Even so, few of us are *certain* that all of these assumptions are true. To handle uncertainty regarding these normative questions about what matters, maximizing expected prudential value is analogous to maximizing expected moral value in decisions under moral uncertainty.[65]

But we also have descriptive uncertainty. The technologies we need in order to implement these approaches have yet to be invented, and it's unclear when they will, if ever. But the two technologies that did not (by themselves) lead to Strong Prudential Longtermism (that is, anti-ageing and cryonics) might still buy us time for uploading or biological uploading to become feasible. Especially anti-ageing seems promising, as anti-ageing research has made some significant advances in recent years.[66] If anti-ageing works, it might raise our life expectancy by several hundred years. This should, then, give us time to perfect either uploading or bio-

---

[65] See Lockhart 2000: 82. Maximizing expected prudential value is open to the same worry about intertheoretic comparisons of value. See Ross 2006: 761–5 and Gustafsson and Torpman 2014: 160–5. Some alternative approaches avoid intertheoretic comparisons of value, for example: My Favourite Theory (Gracely 1996: 331 and Gustafsson and Torpman 2014: 167–70), My Favourite Option (Lockhart 1992: 35–6), the Borda Rule (MacAskill 2016: 989 and MacAskill et al. 2020: 73). But these alternative approaches will, in some cases of predicted future moral progress, lower the expected moral value conditional on every moral theory in which we have any credence. See Gustafsson 2022b: 452–66. So it seems that we have to, as well as we can, rely on intertheoretic comparisons of value. (Still, just like average and total utilitarianism lack a common unit, the average and total views for aggregation in fission cases and within life-paths also lack a common unit. See Broome 2012: 185.)

[66] See de Grey and Rae 2007: 49–308 and Partridge et al. 2020.

logical uploading.[67]

[67] One worry, however, is a prudential analogue of *the Doomsday Argument* — a notorious argument that, taking ourself to be a random sample from all people in history, we should, given our relatively early position, lower our credence in that humanity will colonize the stars and create an enormous number of future people. (See Leslie 1989: 10; 1996: 187–236 and Bostrom 2002: 89–108.) What we may call *the Prudential Doomsday Argument* is an argument that you won't live for an extremely long time. (See Korb and Oliver 1998: 405n2, and, for a similar argument against the likelihood of an eternal afterlife, see Leslie 2008: 520–4 and Page 2010: 397–401.) If your life will be extremely long (or split into an enormous amount of uploads), then most of your observer-moments will be observer-moments where you are much older than you are now, or they will be simulated observer-moments. We apply

> *The Strong Self-Sampling Assumption*   One should reason as if one's present observer-moment was a random sample from the set of all observer-moments in its reference class.

(Bostrom 2002: 126.) We take the reference class for your current observer-moment to include all your observer-moments. So, if you regard your current observer-moment as a random sample from all of your observer-moments, it would be surprising if you got an observer-moment where you are still this young and not a simulation. So, the argument goes, you should consider it unlikely that you will live for an extremely long time or split into an enormous amount of simulations (assuming that you can tell whether you are simulated). Or, at least, you should regard this possibility as less likely than you did before you considered the Prudential Doomsday Argument. Bostrom (2002: 111–5) objects that the relevant reference class should include not only your observer-moments but also all other observer-moments. If so, he argues, the Prudential Doomsday Argument falls apart because, once we take into account that long lives include more observer-moments, we neutralize the adjustment for finding that your current observer-moment is early. But his solution assumes that we already know the average lifespan of the people in our reference class. In our discussion of the feasibility of extreme life extension, this isn't something we know in advance. Moreover, a standard defence against the Doomsday Argument is to adopt

> *The Self-Indication Assumption*   Given the fact that you exist, you should (other things equal) favour hypotheses according to which many observers exist over hypotheses on which few observers exist.

(Bostrom 2002: 66.) This principle neutralizes the Doomsday Argument. (See Bostrom 2002: 122–3.) The trouble is that this kind of move does not seem very plausible against the Prudential Doomsday Argument. The analogue of the Self-Indication Assumption for observer-moments would be

> *The Strong Self-Indication Assumption*   Given the fact that you have a current observer-moment, you should (other things being equal) favour hypotheses according to which many observer-moments exist over hypotheses on which few observer-moments exist.

But this principle no more favours hypotheses with lots of long lives than hypotheses with the same number of observer-moments but with only short lives.

If either uploading or biological uploading becomes technologically feasible and our assumptions about what matters are correct, then uploading could create an enormous amount of prudential value through longevity, fission, and increasing quality of life. So, even if there were only a small chance that these technologies will work during the lifetimes of some currently existing people and we are uncertain whether our assumptions about what matters are correct, we should still get that Strong Prudential Longtermism holds for some currently existing people in terms of their overall expectation of prudential value.[68] This is fully consistent with these technologies being unlikely to work.

6. Longtermism based on Prudential Longtermism

Given Prudential Longtermism, a large number of theories that otherwise wouldn't lead to (impersonal) Longtermism may turn out to do so.[69] Person-affecting views on which we should minimize the strongest complaint would lead to Longtermism.[70] This is so, since the strongest complaints will come from people for whom Prudential Longtermism is true. Likewise, common-sense morality, on which one should prioritize one's family and friends, would lead to Longtermism if Prudential Longtermism holds for a sufficient number of one's family and friends. Self-interest theories would lead to Longtermism if Prudential Longtermism holds for the agent. Finally, person-affecting utilitarianism would lead to Longtermism if Prudential Longtermism holds for a sufficient number of currently existing people.[71]

The practical implications of Longtermism based on Prudential Longtermism would differ in some respects from those of Longtermism based

---

[68] Maximizing expected prudential value may seem to lead to a kind of fanaticism in these kinds of cases where the overall calculation is dominated by a very unlikely but enormously valuable outcome. See Smith 2014, Monton 2019, and Kosonen 2022: 137–239. But deviations from expected utility theory are vulnerable to money pumps. See Gustafsson 2022a and Kosonen 2022: 196–239.

[69] Prudential longtermism is fairly implausible for non-human animals. So the case for Longtermism based on Prudential Longtermism may be weaker on views where non-human animals typically dominate the overall calculation of value. But, once we take Prudential Longtermism into account, it may be that non-human animals no longer dominate.

[70] Parfit's (n.d.: ch. 6) principle 'Minimax Loss: The best outcome is the one in which the greatest loser loses least.'

[71] Bostrom 2003: 311–2.

on total utilitarianism. In addition to prioritizing the reduction of existential risk to safeguard humanity as a whole, Longtermism based on Prudential Longtermism would also prioritize speeding up technological progress in the areas that might help life extension.[72] It would prioritize funding life extension, so that, in the long-run, some of us may still be alive.[73] (Compared to regular Longtermism, Prudential Longtermism would recommend being more willing to take greater existential risks with AI, since fast AI progress plausibly increases the chance of developing life extension in time.) The badness of death plausibly consists, largely, in how much better a person's life would have been in expectation if they had lived on.[74] Consequently, Prudential Longtermism makes avoiding an early death all the more pressing.[75]

References

Aaronson, Scott (2016) 'The Ghost in the Quantum Turing Machine', in S. Barry Cooper and Andrew Hodges, eds., *The Once and Future Turing: Computing the World*, pp. 193–296, Cambridge: Cambridge University Press.

Ahmed, Arif (2014) *Evidence, Decision and Causality*, Cambridge: Cambridge University Press.

— (2020) 'Rationality and Future Discounting', *Topoi* 39 (2): 245–256.

Arias, Elizabeth and Jiaquan Xu (2022) 'United States Life Tables, 2019', *National Vital Statistics Reports* 70 (19): 1–58.

Arrhenius, Gustaf (2000) *Future Generations: A Challenge for Moral Theory*, Ph.D. thesis, Uppsala University.

Arrhenius, Gustaf and Wlodek Rabinowicz (2015) 'The Value of Existence', in Iwao Hirose and Jonas Olson, eds., *The Oxford Handbook of Value Theory*, pp. 424–443, New York: Oxford University Press.

Benford, Gregory et al. (2005) 'Scientists' Open Letter on Cryonics', URL https://www.biostasis.com/scientists-open-letter-on-cryonics/.

Bostrom, Nick (2002) *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, London: Routledge.

---

[72] Bostrom 2003: 313–4.

[73] Compare Keynes's (1923: 80) more pessimistic assessment.

[74] Broome 1993: 83.

— (2003) 'Astronomical Waste: The Opportunity Cost of Delayed Technological Development', *Utilitas* 15 (3): 308–314.

— (2005) 'The Fable of the Dragon-Tyrant', *Journal of Medical Ethics* 31 (5): 273–277.

— (2014) *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press.

Bostrom, Nick and Toby Ord (2006) 'The Reversal Test: Eliminating Status Quo Bias in Applied Ethics', *Ethics* 116 (4): 656–679.

Bostrom, Nick and Rebecca Roache (2007) 'Human Enhancement: Ethical Issues in Human Enhancement', in Jesper Ryberg, Thomas S. Petersen, and Clark Wolf, eds., *New Waves in Applied Ethics*, pp. 120–152, Basingstoke: Palgrave Macmillan.

Brink, David O. (1992) 'Sidgwick and the Rationale for Ratinal Egoism', in Bart Schultz, ed., *Essays on Henry Sidgwick*, pp. 199–240, Cambridge: Cambridge University Press.

Broome, John (1993) 'Goodness Is Reducible to Betterness: The Evil of Death Is the Value of Life', in Peter Koslowski and Yuichi Shionoya, eds., *The Good and the Economical: Ethical Choices in Economics and Management*, pp. 70–84, Berlin: Springer.

— (2004) *Weighing Lives*, Oxford: Oxford University Press.

— (2012) *Climate Matters: Ethics in a Warming World*, New York: Norton.

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

Clark, Andy and David Chalmers (1998) 'The Extended Mind', *Analysis* 58 (1): 7–19.

Crimmins, Eileen M (2015) 'Lifespan and Healthspan: Past, Present, and Promise', *The Gerontologist* 55 (6): 901–911.

Crisp, Roger (1997) *Mill on Utilitarianism*, London: Routledge.

Dainton, Barry (2012) 'On Singularities and Simulations', *Journal of Consciousness Studies* 19 (1–2): 42–85.

Dainton, Barry and Tim Bayne (2005) 'Consciousness as a Guide to Personal Persistence', *Australasian Journal of Philosophy* 83 (4): 549–571.

de Grey, Aubrey and Michael Rae (2007) *Ending Aging: The Rejuvenation Breakthroughs that Could Reverse Human Again in Our Lifetime*, New York: St. Martin's Press.

Doyle, D. John (2018) *What Does it Mean to be Human? Life, Death, Personhood and the Transhumanist Movement*, Berlin: Springer.

Dyson, Freeman J. (1979) 'Time without End: Physics and Biology in an Open Universe', *Reviews of Modern Physics* 51 (3): 447–460.

Gibbard, Allan and William L. Harper (1978) 'Counterfactuals and Two Kinds of Expected Utility', in C. A. Hooker, J. J. Leach, and E. F. Mc-Clennen, eds., *Foundations and Applications of Decision Theory*, vol. I, pp. 125–162, Dordrecht: Reidel.

Gracely, Edward J. (1996) 'On the Noncomparability of Judgments Made by Different Ethical Theories', *Metaphilosophy* 27 (3): 327–332.

Greaves, Hilary (2019) 'Review of Samuel Scheffler, *Why Worry about Future Generations?*', *Ethics* 130 (1): 136–141.

Greaves, Hilary and William MacAskill (forthcoming) 'The Case for Strong Longtermism', in Jacob Barrett, Hilary Greaves, and David Thorstad, eds., *Essays on Longtermism*, Oxford: Oxford University Press.

Gustafsson, Johan E. (2011) 'Phenomenal Continuity and the Bridge Problem', *Philosophia* 39 (2): 289–296.

— (2018) 'The Unimportance of Being Any Future Person', *Philosophical Studies* 175 (3): 745–750.

— (2019) 'Non-Branching Personal Persistence', *Philosophical Studies* 176 (9): 2307–2329.

— (2021) 'Is Psychology What Matters in Survival?', *Australasian Journal of Philosophy* 99 (3): 504–516.

— (2022a) *Money-Pump Arguments*, Cambridge: Cambridge University Press.

— (2022b) 'Second Thoughts about My Favourite Theory', *Pacific Philosophical Quarterly* 103 (3): 448–470.

Gustafsson, Johan E. and Petra Kosonen (forthcoming) 'Do Lefty and Right Matter More than Lefty Alone?', *Erkenntnis*.

Gustafsson, Johan E. and Olle Torpman (2014) 'In Defence of My Favourite Theory', *Pacific Philosophical Quarterly* 95 (2): 159–174.

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

Holtug, Nils (2001) 'The Repugnant Conclusion about Self-Interest', *Danish Yearbook of Philosophy* 36 (1): 49–68.

— (2010) *Persons, Interests, and Justice*, Oxford: Oxford University Press.

Jeffrey, Richard C. (1965) *The Logic of Decision*, New York: McGraw-Hill.

— (1983) *The Logic of Decision*, Chicago: University of Chicago Press, second edn.

Keynes, John Maynard (1923) *A Tract on Monetary Reform*, London: Macmillan.

Korb, Kevin B. and Jonathan J. Oliver (1998) 'A Refutation of the Dooms-

day Argument', *Mind* 107 (426): 403–410.

Kosonen, Petra (2022) *Tiny Probabilities of Vast Value*, Ph.D. thesis, University of Oxford.

Leslie, John (1989) 'Risking the World's End', *Bulletin of the Canadian Nuclear Society* 10 (3): 10–15.

— (1996) *The End of the World: The Science and Ethics of Human Extinction*, London: Routledge.

— (2008) 'Infinitely Long Afterlives and the Doomsday Argument', *Philosophy* 83 (326): 519–524.

Lewis, David (1976) 'Survival and Identity', in Aélie Oksenberg Rorty, ed., *The Identities of Persons*, pp. 17–40, Berkeley: University of California Press.

— (1986) *On the Plurality of Worlds*, Oxford: Blackwell.

Lockhart, Ted (1992) 'Professions, Confidentiality, and Moral Uncertainty', *Professional Ethics* 1 (3–4): 33–52.

— (2000) *Moral Uncertainty and Its Consequences*, New York: Oxford University Press.

MacAskill, William (2016) 'Normative Uncertainty as a Voting Problem', *Mind* 125 (500): 967–1004.

— (2019) "Longtermism", Effective Altruism Blog, URL https://forum.effectivealtruism.org/posts/qZyshHCNkjs3TvSem/longtermism.

MacAskill, William, Krister Bykvist, and Toby Ord (2020) *Moral Uncertainty*, Oxford: Oxford University Press.

McMahan, Jeff (1981) 'Problems of Population Theory', *Ethics* 92 (1): 96–127.

— (2002) *The Ethics of Killing: Problems at the Margin of Life*, Oxford: Oxford University Press.

McTaggart, John McTaggart Ellis (1927) *The Nature of Existence Volume II*, Cambridge: Cambridge University Press.

Merkle, Ralph C. (1992) 'The Technical Feasibility of Cryonics', *Medical Hypotheses* 39 (1): 6–16.

— (1994) 'The Molecular Repair of the Brain, Part I', *Cryonics* 15 (1): 16–31.

Miller, Kevin (2004) 'Cryonics Redux: Is Vitrification a Viable Alternative to Immortality as a Popsicle?', *Skeptic* 11 (1): 24–25.

Minerva, Francesca (2018) *The Ethics of Cryonics: Is It Immoral to Be Immortal?*, Basingstoke: Palgrave Macmillan.

Mogensen, Andreas (forthcoming) 'Would a World without Us Be Worse?', in Jacob Barrett, Hilary Greaves, and David Thorstad, eds., *Essays on Longtermism*, Oxford: Oxford University Press.

Monton, Bradley (2019) 'How to Avoid Maximizing Expected Utility', *Philosophers' Imprint* 19 (18): 1–24.

Nagel, Thomas (1986) *The View from Nowhere*, Oxford: Oxford University Press.

Narveson, Jan (1973) 'Moral Problems of Population', *Monist* 57 (1): 62–86.

Ord, Toby (2020) *The Precipice: Existential Risk and the Future of Humanity*, London: Bloomsbury.

Page, Don N. (2010) 'Scientific and Philosophical Challenges to Theism', in Melville Y. Stewart, ed., *Science and Religion in Dialogue*, pp. 396–410, Oxford: Blackwell.

Parfit, Derek (1971) 'Personal Identity', *The Philosophical Review* 80 (1): 3–27.

— (1984) *Reasons and Persons*, Oxford: Clarendon Press.

— (1986) 'Overpopulation and the Quality of Life', in Peter Singer, ed., *Applied Ethics*, pp. 145–164, Oxford: Oxford University Press.

— (1993) 'The Indeterminacy of Identity: A Reply to Brueckner', *Philosophical Studies* 70 (1): 23–33.

— (1995) 'The Unimportance of Identity', in Henry Harris, ed., *Identity: Essays Based on Herbert Spencer Lectures Given in the University of Oxford*, pp. 13–45, Oxford: Clarendon Press.

— (2017) *On What Matters: Volume Three*, Oxford: Oxford University Press.

— (n.d.) *On Giving Priority to the Worse Off*, unpublished manuscript.

Partridge, Linda, Matias Fuentealba, and Brian K. Kennedy (2020) 'The Quest to Slow Ageing through Drug Discovery', *Nature Reviews Drug Discovery* 19 (8): 513–532.

Perry, John (1972) 'Can the Self Divide?', *The Journal of Philosophy* 69 (16): 463–488.

Ramsey, F. P. (1928) 'A Mathematical Theory of Saving', *The Economic Journal* 38 (152): 534–559.

Rawls, John (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.

— (1999) *A Theory of Justice*, Cambridge, MA: Harvard University Press, revised edn.

Ross, Jacob (2006) 'Rejecting Ethical Deflationism', *Ethics* 116 (4): 742–768.

— (2014) 'Divided We Fall', *Philosophical Perspectives* 28 (1): 222–262.

Sagan, Carl (1994) *Pale Blue Dot: A Vision of the Human Future in Space*, New York: Random House.

Sandberg, Anders and Nick Bostrom (2008) 'Whole Brain Emulation: A Roadmap', Technical Report 2008-3, Future of Humanity Institute, Oxford.

Scheffler, Samuel (2013) *Death and the Afterlife*, New York: Oxford University Press.

— (2018) *Why Worry About Future Generations?*, New York: Oxford University Press.

Shaw, David (2009) 'Cryoethics: Seeking Life After Death', *Bioethics* 23 (9): 515–521.

Shoemaker, Sydney (1970) 'Persons and Their Pasts', *American Philosophical Quarterly* 7 (4): 269–285.

Sidgwick, Henry (1907) *The Methods of Ethics*, London: Macmillan, seventh edn.

Smith, Nicholas J. J. (2014) 'Is Evaluative Compositionality a Requirement of Rationality?', *Mind* 123 (490): 457–502.

Steele, Katie (forthcoming) 'Longtermism and Neutrality about More Lives', in Jacob Barrett, Hilary Greaves, and David Thorstad, eds., *Essays on Longtermism*, Oxford: Oxford University Press.

Tappenden, Paul (2011) 'Expectancy and Rational Action Prior to Personal Fission', *Philosophical Studies* 153 (2): 299–306.

Temkin, Larry S. (1987) 'Intransitivity and the Mere Addition Paradox', *Philosophy & Public Affairs* 16 (2): 138–187.

— (2008) 'Is Living Longer Living Better?', *Journal of Applied Philosophy* 25 (3): 193–210.

— (2012) *Rethinking the Good*, New York: Oxford University Press.

Thomas, Teruji (forthcoming) 'The Asymmetry, Uncertainty, and the Long Term', *Philosophy and Phenomenological Research*.

Unger, Peter (1990) *Identity, Consciousness and Value*, New York: Oxford University Press.

Vita-More, Natasha and Danliel Barranco (2015) 'Persistence of Long-Term Memory in Vitrified and Revived *Caenorhabditis elegans*', *Rejuvenation Research* 18 (5): 458–463.

Williams, Bernard (1973) 'The Makropulos Case: Reflections on the Tedium of Immortality', in *Problems of the Self: Philosophical Papers 1956–1972*, pp. 82–100, Cambridge: Cambridge University Press.