Death and prudential deprivation

Matthew W. G. McClure

1 Introduction

Is death bad for us?[1] The question falls prey to two ambiguities, those between *being dead* and *dying*, and between *existential* and *universal* quantification (Fletcher 2016: 145–6, 148). Here I am concerned only with dying:

> Weak harm thesis: there exist lives where dying is (prudentially) bad for the dier. (Sometimes dying is bad for us.) In symbols: $\exists S \exists t (\llbracket \varnothing \blacktriangleright S \rrbracket_t)$. (For notation see appx. A.)

> Strong harm thesis: for all lives, dying is (prudentially) bad for the dier. (Dying is always bad for us.) In symbols: $\forall S \exists t (\llbracket \varnothing \blacktriangleright S \rrbracket_t)$.

In this paper I'll argue for the weak harm thesis from a principle about which futures are bad for a prudential subject, the *future-comparison principle*. A plausible strengthening of this principle yields an argument against the strong harm thesis, and gives us reason to doubt the 'annihilation' account of the badness of dying (dying is bad in that it destroys the dier).

2 The deprivation account and the future-comparison principle

Why is dying bad for us? On the 'comparative' or 'deprivation' account (DA): Dying (sometimes) prevents us from being the subjects of wellbeing that we would otherwise be, and failing to be the subject of this wellbeing is bad for us (since being so would be better for us). So sometimes dying is bad for us (cf. Fletcher 2016: 148–150; Luper 2009: ch. 5; Benatar 2017: 101–102).

The argument for the weak harm thesis from the prudential value of futures, a species of DA, focuses on one particular life, Σ.

> Σ has a linear uphill distribution, and her prudential level would continue to rise were it not for her death (fig. 1, p. 2). Her death does not feature any ancillary harms like pain.

Σ is of course an abstraction, but there are lives of concrete prudential subjects that share the important features – net positive wellbeing before a death without significant ancillary harm, and plausible net positive wellbeing after the time of (actual) death.

The following principle about wellbeing and futures is the basis of my argument.

> Future-comparison principle (FCP): A sufficient condition on h being bad for S at t is for there to exist a t-accessible future g such that S's wellbeing in g exceeds that in h. In symbols: $\llbracket h \blacktriangleright S \rrbracket_t \Leftarrow \exists g (tRg \wedge w(S)_g > w(S)_h)$.

---

[1] I talk of *harming* and being *bad for* interchangeably, and of the *prudential* as having to do with wellbeing.
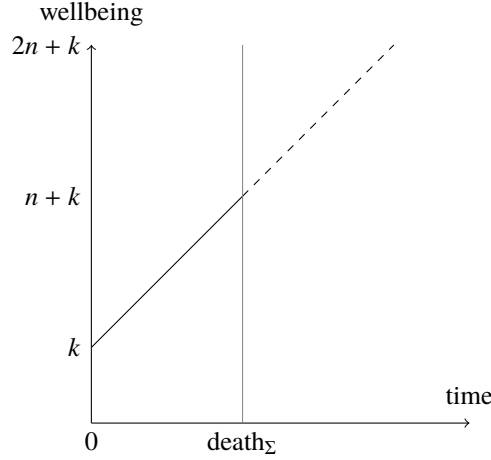
Figure 1: Σ's life. Dying is bad for her.

I take it that FCP should be relatively uncontroversial; prima facie, it seems true: surely something is bad for you if it makes you worse-off in the long run. An argument for the weak harm principle is obtained when a characterisation of dying is given.

> NF: Dying is the null future ($\varnothing$); it is therefore of no prudential value: $\forall S(w(S)_\varnothing = 0)$

The argument comes quite easily. Let G be any death$_\Sigma$-accessible future with positive wellbeing (such things exist by definition – see the dashed line in fig. 1). By NF, dying ($\varnothing$) has prudential value 0. So G is accessible at death$_\Sigma$ and has greater wellbeing than $\varnothing$. But for $\varnothing$ to be bad for Σ at death$_\Sigma$, it's sufficient that there be better death$_\Sigma$-accessible futures for Σ (by FCP): viz., G. So $\varnothing$ is bad for Σ at death$_\Sigma$. (For the technical details, see appx. B.)

Since we define $\blacktriangleright$ using R (the accessibility relation), it's worth pausing for a moment and considering what natural constraints on the accessibility relation yield for the badness of dying.

Transitivity is a fairly natural constraint for temporal reasoning (especially if we think of R as a 'precedence' relation as in tense logics), though it would require a slight tweak of our reading of 'accessibility' (e.g. from 'directly precedes' to 'precedes'). Consider S at t with an accessible future $u, \ldots, \varnothing$ ($uRu^*, \ldots, u^{*\cdots*}R\varnothing$). Then $tRu, tRu^*, \ldots, tR\varnothing, tR\langle u, u^*\rangle, \ldots, tR\langle u^*, u^{**}\rangle, \ldots, tR\langle u, u^*, u^{**}\rangle, \ldots$: the t-accessibility of any future of which any part has net positive wellbeing entails the badness of death at t.

Reflexivity is a less natural constraint (though it makes more sense if we think of precedence as $\leq$ rather than $<$), but it yields an interesting result. Consider some moment t at which $\varnothing$ is accessible; by reflexivity, tRt, so we can recursively construct an infinity of futures $\varnothing$; $t, \varnothing$; $t, t, \varnothing$; $t, \ldots, \varnothing$. If t has net positive prudential value, $\varnothing$ is bad for the prudential subject at t.

2

3 Objections

An influential objection to the badness of death comes from Epicurus:

> Thus he is a fool who says he fears death not because it will be painful while present but because it is painful when it is still to come. For that which while present causes no distress causes unnecessary pain when merely anticipated. So death, the most frightening of bad things, is nothing to us; since when we exist, death is not yet present, and when death is present, then we do not exist. (Epicurus 1990: 315)

Following Bradley (2004: 1), I put this 'timing' challenge in the form of a modus tollens inference:

1. for any f, if f is bad for S, it is bad for her at some time

2. there is no time at which dying is bad for S

3. so dying is not bad for S

The natural response is to claim that (2) is false, since FCP is concerned with badness at particular instants. Keeping this, and that dying is a future, in mind, it makes perfect sense to say there are times at which dying is bad for S, since dying is a future accessible at particular times and manifest at particular times. FCP entails that the prudential disvalue is borne by the instant before death, since in this instant dying is an accessible future, and the prudential value of an instant is partially determined by the future it manifests: to be stuck between Scylla and Charybdis is bad for S even before one of them eats her.

The classical 'symmetry' challenge of Lucretius (1997: 96–7 (3.972–979)) may be addressable too.

> Look back upon the ages of time past
> Eternal, before we were born, and see
> That they have been nothing to us, nothing at all.
> This is the mirror nature holds for us
> To show the face of time to come, when we
> At last are dead. Is there in this for us
> Anything horrible? Is there anything sad?
> Is it not more free from care than any sleep?

An argument for the nonbadness of dying can be given when we formulate the 'symmetry' principle:

> SP: Coming into existence (birth, the null past) and going out of existence (dying, the null future) are alike in all relevant respects.

As there is no question of coming into existence being prudentially good or bad for us, since there was no 'us', so, by SP, there is no question of going out of existence being prudentially good or bad for us, since there will be no 'us'. Imagine one flipped the axis
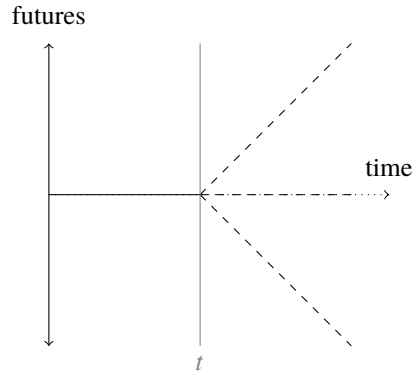
futures

time

t

Figure 2: Asymmetry between the past and future

of time (by 'the mirror of nature'): birth$_{\text{flipped}}$ would be dying$_{\text{not flipped}}$, as in F. Scott Fitzgerald's 'The curious case of Benjamin Button'. But wouldn't it then fail to be *bad* for us? By SP the flipped and non-flipped lives are alike in all relevant respects, so dying$_{\text{not flipped}}$ is not bad for us.[2]

FCP allows us to assert, and to explain, the asymmetry: At most one past is accessible to any instant constrained by a history, so FCP does not apply to coming into existence. Prudential asymmetry mirrors temporal asymmetry (fig. 2, p. 4).

So much for the classical worries: let's consider an interesting objection to FCP. Consider two accessible futures – f, which has (very high) n prudens[3], and g, which has $n - \varepsilon$ (marginally less). Each would vastly improve S's level of wellbeing, but on FCP g is *bad* for S. But this is absurd.

The response has two parts. First, we bite the bullet: g is bad for S given the accessibility of f. Then we show that biting the bullet is not so bad: g may be bad, but it's not as bad as e.g. a future with $n - 35$ prudens, or $\varnothing$. Value comes, to some extent, in degrees – the goodness/badness of a future is proportional to its prudential value. This has the very intuitive implication that how bad dying is for S is proportional to the net value of her other accessible future.[4] It is also a natural bedfellow to FCP. (Turns out the bullet is not so bad after all.)

As an example, take Fletcher's (2016: 149) 'dastardly David' case, in which a senior

---

[2]A formulation a little closer to the poem (Lucretius seems to be concerned only with the nonbadness of being dead) is given by Luper (2009: 61–2) (his additions in parenthesis):

1. It is not bad for us that we once failed to exist.

2. Our posthumous nonexistence is like our prevital nonexistence in all relevant respects.

3. If two things are alike in all relevant respects, and one of them is not bad for us, then the second is not bad for us either.

4. So it is not bad for us that we will fail to exist once more.

5. (If the significant effect of a process is to place us into a state that is not bad for us, then that process is not bad for us either.)

6. (Therefore the process of death is not bad for us.)

[3]A pruden is a unit of prudential value.

[4]Where there are more than one, some weighting is required, but this is outside of this paper's scope.

academic (David) prevents a brilliant junior academic (Beth) from taking up a prestigious and well-funded research fellowship so she can teach a course for him. Her wellbeing is still high – she still has a satisfying academic career, &c. – but it would've been higher had she been able to take up the fellowship. It should be clear that being prevented from taking up the fellowship is bad for Beth, even if her wellbeing is still very high. Though it is of course not as bad for her as, e.g., dying in a bicycle accident on her way to work.

4 Is death *always* bad?

So, dying is bad for some prudential subjects – but is it bad for all? Not if we accept the *strengthened future-comparison principle*:

> FCP+: A *necessary* and sufficient condition on h being bad for S at t is for there to exist a t-accessible future g such that S's wellbeing in g exceeds that in h. In symbols: $[\![h \blacktriangleright S]\!]_t \Leftrightarrow \exists g(tRg \wedge w(S)_g > w(S)_h)$.

To see the plausibility of FCP+, consider that whenever something is bad for you, your level of wellbeing falls relative to the best alternative. Indeed, it is bad for you only when your level of wellbeing is lower than the best alternative.[5] Thus that which is bad is extensionally equivalent with that which lowers your level of wellbeing relative to the alternatives, which is what FCP+ claims.
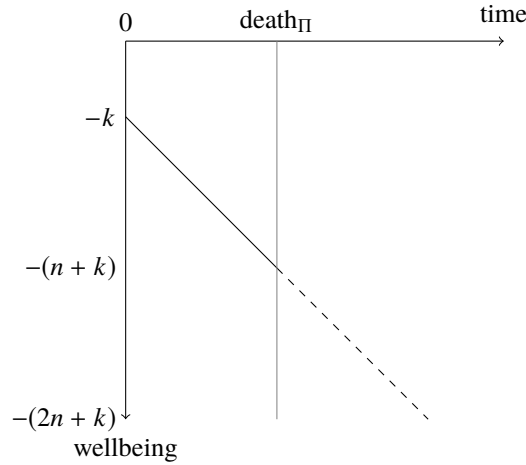

Figure 3: Π's life. Not dying would be bad for her.

The argument against the strong harm thesis considers Π, who is the inverse of Σ (fig. 3, p. 5). At death_Π there are no accessible futures at which her wellbeing exceeds that in $\varnothing$, so by FCP+ it is not bad for her. But for that distinct future h, in which she does not die, there exists some future, $\varnothing$, whose prudential value exceeds that of h. To die would not be bad for Π; to not die would be bad for Π.

---

[5]Note alternatives are given by prudential accessibility, which needn't be constrained by nomological possibility – or even metaphysical possibility. To evaluate alternatives, given the right theory of wellbeing, just evaluate the counterfactual.

It is, however, worth noting that, practically speaking, death will always feature ancillary harms – pain, fear, projects being left unfinished. These are bad for S when S dies, but death in itself is only sometimes bad for her.

5 The supposed badness of ceasing to be

In the foregoing, I have argued that the badness of dying is not a special type of harm, but merely a token of comparative harm (the badness of death is not peculiar to death). Benatar (2017: 102–110) suggests, in addition to DA, that dying is a special harm, bad for S in that it destroys her; it makes her cease to be (the 'annihilation' account). I will briefly suggest a reason to doubt this. (It is inconsistent with FCP+.)

Dying is not bad in virtue of annihilation. If one could destroy a life that would have equal wellbeing to as it would were it not destroyed, death would be no worse for its subject. Λ (fig. 4, p. 6) is an example, and so is ID.
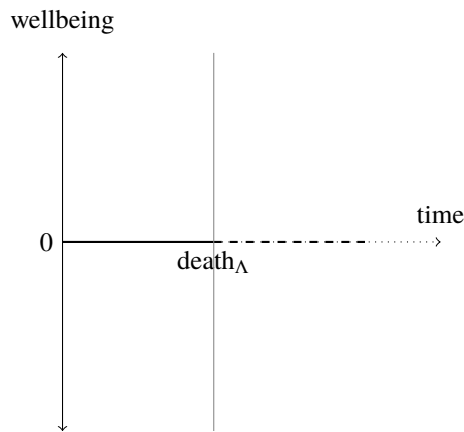

Figure 4: Λ's life. Dying doesn't seem to be bad for her in virtue of her ceasing to be.

> Instantaneous death (ID): a subject dies, but only for an instant. She does not remain dead, and continues to live a life of relatively high wellbeing.

Similar concrete cases have occurred, though they usually feature ancillary harms and last longer than an instant. In ID cases it hardly seems right to say that annihilation is bad for S since it makes the badness of death for S independent from that which is prudentially tangible.[6] What, one might ask, is so special about death?

Annihilation may not be in itself bad, as per the counterexamples[7], but it is very often associated with harms – ancillary harms, like pain; and deprivation harms of the sort considered above: that one will cease to be entails that one will not have access to other futures, which might be better.

---

[6] It fails to satisfy the experience requirement, for example.

[7] That ID is impossible given the permanence of death is not relevant – annihilation and mortal harm still come apart conceptually (recall fn 5).

Perhaps we can explain the annihilation intuition, then, by saying: dying entails that one can not manifest alternative futures, and the badness of death (when it is bad) consists in not being able to manifest better alternative futures than death. Destroying *Guernica* is bad because (and insofar as) it prevents those goods that would've obtained had *Guernica* persisted from obtaining. When death is bad, its badness is guaranteed by annihilation, but it does not consist in annihilation.

6 Conclusion

Where does the discussion leave us? It seems the future-comparison principle (with the null future characterisation of dying) gives us good reason to accept the weak harm thesis on grounds of comparative harm/prudential deprivation. FCP can address influential objections to the weak harm thesis from timing and symmetry, and seems to address the objection regarding marginal differences in wellbeing. Further, the strengthened future-comparison principle gives us good reason to reject the strong harm thesis on comparative grounds, and explains why the annihilation account fails despite the annihilation intuition not being misplaced.

And the upshot of all of this for us concrete prudential subjects? Dying is a harm when and only when things could be in some sense better than death.

Some parenthetical remarks of a technical character to finish with: The FCP(+) framework seems a useful tool for prudential analysis in general due to its enabling us to be (i) more abstract in its almost 'metaprudential' character (independence of any particular theory), and (ii) more precise and explicit in its formal framework (analogous to modal logic). It may be found helpful when applied to various other issues concerning wellbeing, e.g. those of interpersonal wellbeing comparison and the prudential value of the shape (structural features) of a life. FCP(+) raises its own questions, two of the most significant of which being (i) what constraints there are on prudential accessibility (transitive? reflexive? subset of nomological, metaphysical, &c. accessibility relation?), and (ii) whether the use of multiple prudential valuation functions, multimodal extensions, &c. allow for richer analysis (e.g. an acceptability relation to model the prudential correlate of obligatoriness/impermissibility).

References

Benatar, D. (2017). *The human predicament: A candid guide to life's biggest questions.* Oxford.

Bradley, B. (2004). When is death bad for the one who dies? *Noûs, 38*(1), 1–28.

Cahn, S. M. (Ed.). (1990). *Classics of western philosophy* (3rd ed.). Hackett.

Epicurus. (1990). Letter to Menoeceus (B. Inwood & L. P. Gerson, Trans.). In S. M. Cahn (Ed.), *Classics of western philosophy* (3rd ed.). Hackett.

Fletcher, G. (2016). *The philosophy of well-being: An introduction.* Routledge.

Lucretius. (1997). *On the nature of the universe* (R. Melville, D. Fowler, & P. Fowler, Eds.). Oxford.

Luper, S. (2009). *The philosophy of death.* Cambridge.

A Notation

- $S, T, \ldots$ are variables in domain $\mathcal{L}$ (prudential subjects)

- $t, u, \ldots$ are variables in domain $\mathcal{T}$ (instants) in the metalanguage

- $f, g, \ldots$ are variables (in $\mathcal{F}$) standing for futures, which are chains of instants (sets of instant-pairs) along the accessibility relation ($tRu, uRv, \ldots$) (metalanguage)

- $\Sigma, \ldots (\in \mathcal{L})$ and $\varnothing (\in \mathcal{F})$ are constants

- $\blacktriangleright \subseteq (\mathcal{F} \times \mathcal{L})$ – it's read '(future) is bad for (prudential subject)'

- w is the prudential valuation function (to an ordered set like $\mathbf{R}$), at instants and over periods of time

- $[\![\cdot]\!].$ is the assignment function (at instants)

B Technical details of the argument

1. $G := \imath g(\text{death}_\Sigma RG \wedge w(\Sigma)_G > 0)$ <div style="float:right">(def.)</div>

2. $\forall S(w(S)_\varnothing = 0)$ <div style="float:right">(NF)</div>

3. $w(\Sigma)_\varnothing = 0$ <div style="float:right">($\forall$-elim. 2)</div>

4. $[\![h \blacktriangleright S]\!]_t \Leftarrow \exists g(tRg \wedge w(S)_g > w(S)_h)$ <div style="float:right">(FCP)</div>

5. $[\![\varnothing \blacktriangleright S]\!]_t \Leftarrow \exists g(tRg \wedge w(S)_g > 0)$ <div style="float:right">($\forall$-elim. 4)</div>

6. $[\![\varnothing \blacktriangleright \Sigma]\!]_t \Leftarrow \exists g(tRg \wedge w(\Sigma)_g > 0)$ <div style="float:right">($\forall$-elim. 5)</div>

7. $[\![\varnothing \blacktriangleright \Sigma]\!]_{\text{death}_\Sigma} \Leftarrow \exists g(\text{death}_\Sigma Rg \wedge w(\Sigma)_g > 0)$ <div style="float:right">($\forall$-elim. 6)</div>

8. $\text{death}_\Sigma RG \wedge w(\Sigma)_G > 0$ <div style="float:right">(G def.)</div>

9. $\exists g(\text{death}_\Sigma Rg \wedge w(\Sigma)_g > 0)$ <div style="float:right">($\exists$-in. 8)</div>

10. $[\![\varnothing \blacktriangleright \Sigma]\!]_t$ <div style="float:right">(m.p. 7, 9)</div>

That G denotes should be easy to see from the stipulation of Σ (fig. 1, p. 2). (5–7) are valid instances of universal instantiation since h, S, and t are free.