Death and Prudential Deprivation

MATTHEW W. G. MCCLURE

Abstract

Dying is (sometimes) bad for the dier because it prevents her from being the subject of wellbeing she otherwise would (the *deprivation account*). I argue for this from a (plausible) principle about which futures are bad for a prudential subject (the *future-comparison principle*). I explain how this principle rules out two influential objections to the badness of death (*timing* and *symmetry*), and defend it from objections. A *strengthened future-comparison principle* yields that death is not always bad, and that the badness of death does not consist in that it destroys the dier.

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. This narrows things down to claims about quantification:

> *Weak harm thesis*: There exist lives where dying is (prudentially) bad for the dier. (Sometimes dying is bad for us.) In symbols: $\exists S \exists t([\![\varnothing \blacktriangleright S]\!]_t)$. (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: $\forall S \exists t([\![\varnothing \blacktriangleright S]\!]_t)$.

In this paper I'll argue for the weak harm thesis from a principle concerning 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).

[1] I use 'harming' and 'bad for' interchangeably, and use 'prudential', 'prudentially', &c., as having to do with wellbeing.

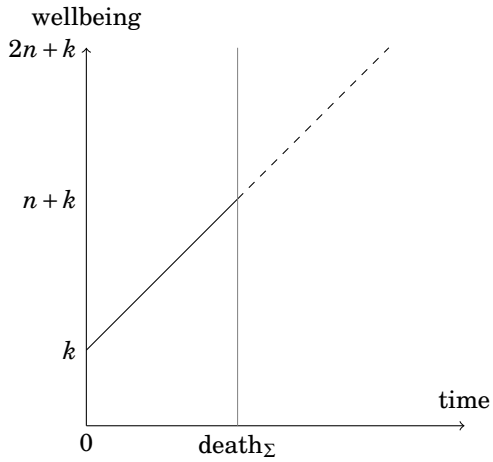

Figure 1: Σ's life. Dying is bad for her.

2 *The deprivation account and the future-comparison principle*

Why is dying bad for us? On the '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 dying is sometimes 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: Σ has a linear uphill distribution, and her prudential level would continue to rise were it not for her death (fig. 1, p. 30). 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 its important features: net positive wellbeing before a death without significant ancillary harm, and plausible net positive wellbeing after the time of (actual) death.

To evaluate Σ, we formulate a framework modelling wellbeing over time, the 'comparative' framework. We start off with an ordinary Kripke triple often seen in the semantics of nonclassical logics

$$\langle \mathscr{T}, R, [\![\cdot]\!]_{\cdot} \rangle$$

interpreted temporally, with $\mathscr{T}$ being a set of instants and R a binary relation on $\mathscr{T}$, interpreting tRu as u being prudentially accessible from t. $[\![\cdot]\!]_{\cdot}$ is the denotation function at instants; $[\![p]\!]_t$ iff p is true at t. In addition to instants, we have futures (this set $\mathscr{F}$), tuples of instants along the accessibility relation; if $\langle u, u^*, u^{**}, u^{***}, \ldots\rangle$ is a future, then $uRu^* \wedge u^*Ru^{**} \wedge u^{**}Ru^{***} \wedge \ldots$.[2] We impose a constraint on R similar to seriality ($\forall t \exists u(tRu)$); specifically, future seriality ($\forall t \exists f(tRf)$): all instants access a future (not necessarily an instant).

This all corresponds to the metalanguage of, say, a modal logic. We introduce further notions in a form corresponding to object language operators:

- a function w which maps prudential subjects to the reals (any sufficiently large totally ordered set will do) at instants and futures, interpreting $w(S)_t$ as S's wellbeing at t and $w(S)_f$ her wellbeing in f.
- a relation $\blacktriangleright$ between futures and prudential subjects, interpreting $[\![f \blacktriangleright S]\!]_t$ as f being prudentially bad for S at t.

Given all this, the following principle about wellbeing and futures is the basis of a deprivation-style argument for the weak harm thesis:

> *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: $[\![h \blacktriangleright S]\!]_t \Leftarrow \exists g(tRg \wedge w(S)_g > w(S)_h)$.

It seems 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. The argument for the weak harm principle is obtained when a characterisation of dying is given:

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

The argument, then, 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$_\Sigma$. The weak harm thesis follows immediately by existential generalisation. (For the technical details, see appx. B.)

[2] Naturally, if $\langle \ldots, t\rangle$ and $\langle u, \ldots\rangle$ are futures, then $\langle \ldots, t\rangle R \langle u, \ldots\rangle$ iff tRu.

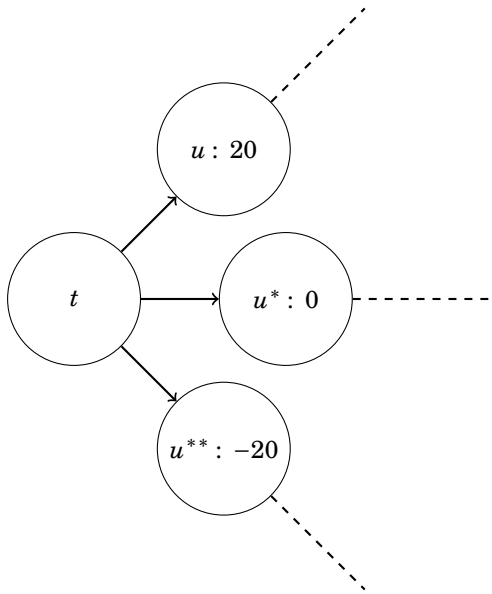

Figure 2: On FCP, $[\![\langle u^* \rangle \blacktriangleright S]\!]_t$ and $[\![\langle u^{**} \rangle \blacktriangleright S]\!]_t$ (since tRu and $w(S)_u > w(S)_{u^*}$ and $w(S)_u > w(S)_{u^{**}}$).

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

Transitivity ($tRu \wedge uRv \implies tRv$) 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 $u,\ldots,\varnothing$ ($uRu^*,\ldots,u^{*\cdots*}R\varnothing$). It then follows that

$$tRu, tRu^*, \ldots, tR\varnothing, tR\langle u, u^*\rangle, \ldots, tR\langle u^*, u^{**}\rangle, \ldots, tR\langle u, u^*, u^{**}\rangle, \ldots$$

and, consequently, that the t-accessibility of any future of which any part has net positive wellbeing entails the badness of death at t (since this positive part is accessible at t, and of greater prudential value than $\varnothing$).

Reflexivity (tRt) is a less natural constraint, 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

$$\varnothing; \quad t,\varnothing; \quad t,t,\varnothing; \quad t,\ldots,\varnothing$$

It follows trivially that if t has net positive prudential value for S, $\varnothing$ is bad for S at t.

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 the fact 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:

> *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.

There is no question of coming into existence being prudentially good or bad for us, since there was no 'us'. Correspondingly, 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 if one flipped the axis of time (by 'the mirror of nature'): $\text{birth}_{\text{flipped}}$ would be $\text{dying}_{\text{not flipped}}$, as in F. Scott Fitzgerald's 'The Curious Case of Benjamin Button'. But wouldn't $\text{dying}_{\text{not flipped}}$ then fail to be *bad* for us? By SP the flipped and non-flipped lives are alike in all relevant respects, so $\text{dying}_{\text{not flipped}}$ is not bad for us.[3]

[3] 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

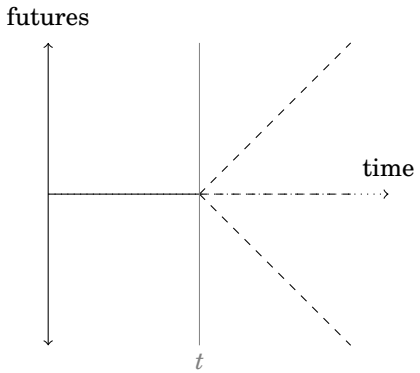

Figure 3: Asymmetry between the past and future

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. 3, p. 35).

So much for the classical worries: let's consider objections to FCP specifically. The two particular objections concern a supposed consequentialist commitment FCP entails, and cases of marginal difference in wellbeing between futures, respectively.

The consequentialism objection proposes a counterexample to FCP. Consider a subject S who is having an awful life, her wellbeing significantly below 0 ('worse than nothing'). S is starving to death and strongly desires to be given food, even thought she knows that, if saved, things will remain very bad.[4] On FCP, the objection goes, we ought not give S food: we ought to let her starve, or perhaps euthanise her.

But this objection is misplaced for two reasons. Firstly, whether FCP will judge death bad for S in this case turns on first-order prudential considerations: if desire-satisfaction contributes significantly to wellbeing, and S's will to live is strong enough, it is plausible that the satisfaction of this desire (giving her food)

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.)

[4]This example is due to Max Klaassen.

will bring her wellbeing up to above 0. Secondly, the comparative framework and FCP are both purely axiological, concerned only with when dying is *bad* for a subject. It says nothing (normative) about what we're obligated to do given these axiological facts, and is therefore not committed to the claim that if not dying is prudentially bad for S, S ought die. So even if S's wellbeing would remain negative after surviving, it doesn't follow that S ought die. If we accept utilitarianism, it follows that S ought die, but the FCP doesn't commit one to utilitarianism. It appears the consequentialism objection itself presupposes consequentialism.[5]

The objection concerning marginal differences in wellbeing runs as follows. Consider two accessible futures: f, which has (very high) n prudens,[6] 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, for example, 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 the degree to which dying is bad for S is proportional to the net value of her other accessible future.[7] It also sits very nicely with 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 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, for example, 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 following principle:

> *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-

[5] As an interesting technical point, FCP+ is structurally similar to consequentialism; interprting ► as an impermissibility relation and w as the theory's axiological function yields consequentialism.

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

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

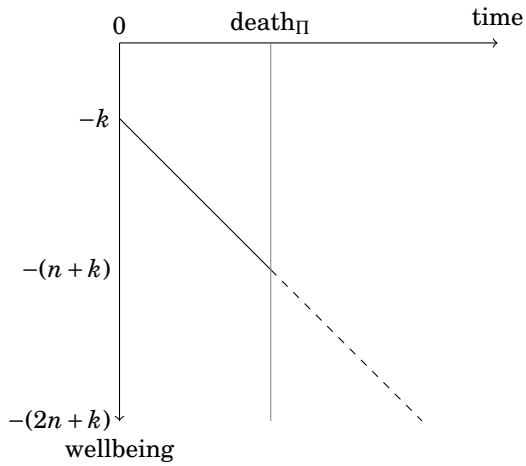

Figure 4: Π's life. Not dying would be bad for her.

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.[8] 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.

The argument against the strong harm thesis considers Π, who is the inverse of Σ (fig. 4, p. 37). At death$_\Pi$ 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.

Practically speaking, however, it's worth noting that 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.

[8]Note that 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.

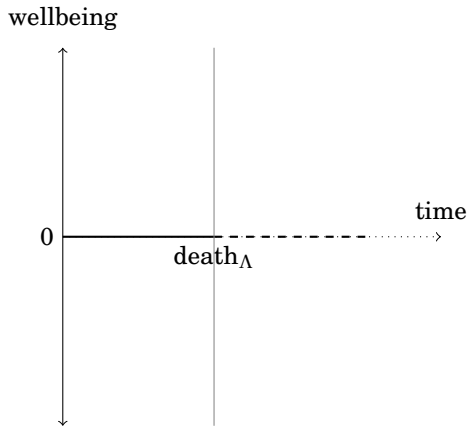

Figure 5: Λ's life. Dying doesn't seem to be bad for her in virtue of her ceasing to be.

5 *The supposed badness of ceasing to be*

In the discussion above, I have argued that the badness of dying is not a special type of harm, but merely a token of comparative harm, that the badness of death is not peculiar to death. Benatar (2017: 102–110) suggests, in addition to DA, that dying is a special harm. The 'annihilation' account says that dying is bad for S in that it destroys her, makes her cease to be. This is inconsistent with FCP+, but we have independent reason to doubt it.

Dying is not bad in virtue of annihilation. Consider some life that, if destroyed, would have the same level wellbeing as it would otherwise have; as examples, take Λ (fig. 5, p. 38) and instantaneous death (below). Were such a life destroyed, death would be no worse from the perspective of its subject; wellbeing would be, broadly speaking, unaffected.

> *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.[9] (The same goes for Λ.) What, one might

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

ask, makes death such a special sort of harm?

Annihilation may not be bad in itself, as per the counterexamples,[10] but it is very often associated with harms. These include ancillary harms, like pain, and comparative harms of the sort considered above – that one will cease to be entails that one will not have access to other futures in which one's wellbeing might be higher.

Perhaps we can explain the annihilation intuition, then, by saying that 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. For example, there are certain goods which obtain because the painting *Guernica* exists. Destroying *Guernica* is bad because (and insofar as) it prevents those goods from obtaining. When death is bad, its badness is guaranteed by annihilation, but does not consist in annihilation.

6 Conclusion

Where does this 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 objections levelled against it. 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. Plausibly, death is bad for just about everyone.

Some parenthetical remarks of a technical character to finish with: The comparative 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, for example, those of interpersonal wellbeing comparison and the prudential value of the shape (structural features) of a life. The comparative framework raises its own questions, two of the most significant of which being (i) what constraints there are on prudential accessibility, and (ii) whether the use of multiple prudential valuation functions, multimodal

[10] That ID is impossible given the permanence of death is not relevant – annihilation and mortal harm are still conceptually independent (recall fn 8).

extensions, &c. allow for richer analysis (e.g. an acceptability relation to model the prudential correlate of obligatoriness/impermissibility).

Acknowledgements

For feedback on this paper I am grateful to an audience at St. Andrews; particularly helpful were the comments of Max Klaassen, Matthew Green, and Andrew Fish. Thanks are due also to Annika Cleland-Hura for helpful editorial suggestions.

References

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

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

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

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

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

A Notation

- $S, T, \ldots$ are variables in domain $\mathscr{L}$ (prudential subjects)
- $t, u, \ldots$ are variables in domain $\mathscr{T}$ (instants) in the metalanguage
- $f, g, \ldots$ are variables (in $\mathscr{F}$) standing for futures, which are chains of instants (tuples) along the accessibility relation ($tRu, uRv, \ldots$) (metalanguage)
- $\Sigma, \ldots (\in \mathscr{L})$ and $\varnothing (\in \mathscr{F})$ are constants
- $\blacktriangleright \subseteq (\mathscr{F} \times \mathscr{L})$ – 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
- $[\![\cdot]\!]_{\cdot}$ is the assignment function (at instants)

B The argument for the weak harm thesis

1. $G := \imath g(\mathrm{death}_\Sigma R G \wedge w(\Sigma)_G > 0)$ (def.)
2. $\forall S(w(S)_\varnothing = 0)$ (NF)
3. $w(\Sigma)_\varnothing = 0$ ($\forall$-elim. 2)
4. $[\![h \blacktriangleright S]\!]_t \Leftarrow \exists g(tRg \wedge w(S)_g > w(S)_h)$ (FCP)
5. $[\![\varnothing \blacktriangleright S]\!]_t \Leftarrow \exists g(tRg \wedge w(S)_g > 0)$ ($\forall$-elim. 4)
6. $[\![\varnothing \blacktriangleright \Sigma]\!]_t \Leftarrow \exists g(tRg \wedge w(\Sigma)_g > 0)$ ($\forall$-elim. 5)
7. $[\![\varnothing \blacktriangleright \Sigma]\!]_{\mathrm{death}_\Sigma} \Leftarrow \exists g(\mathrm{death}_\Sigma Rg \wedge w(\Sigma)_g > 0)$ ($\forall$-elim. 6)
8. $\mathrm{death}_\Sigma RG \wedge w(\Sigma)_G > 0$ (G def.)
9. $\exists g(\mathrm{death}_\Sigma Rg \wedge w(\Sigma)_g > 0)$ ($\exists$-in. 8)
10. $[\![\varnothing \blacktriangleright \Sigma]\!]_t$ ($\rightarrow$-elim. 7, 9)
11. $\exists t([\![\varnothing \blacktriangleright \Sigma]\!]_t)$ ($\exists$-in. 10)
12. $\exists S \exists t([\![\varnothing \blacktriangleright S]\!]_t)$ ($\exists$-in. 11)

That G denotes should be easy to see from the stipulation of Σ (fig. 1, p. 30). (5–7) are valid instances of universal instantiation since h, S, and t are free.