

Critical-set views, biographical identity, and the long term

Elliott Thornley (Global Priorities Institute, University
of Oxford)

Global Priorities Institute | March 2024

GPI Working Paper No. 7-2024

Please cite this working paper as: Thornley, E. Critical-set views, biographical identity, and the long term. *Global Priorities Institute Working Paper Series, No. 7-2024*. Available at: <https://globalprioritiesinstitute.org/critical-set-views-biographical-identity-and-the-long-term-thornley>



Critical-Set Views, Biographical Identity, and the Long Term

*Elliott Thornley**

Forthcoming in the *Australasian Journal of Philosophy*

Abstract: Critical-set views avoid the Repugnant Conclusion by subtracting some constant from the welfare score of *each life* in a population. These views are thus sensitive to facts about biographical identity: identity between lives. In this paper, I argue that questions of biographical identity give us reason to reject critical-set views and embrace the total view. I end with a practical implication. If we shift our credences towards the total view, we should also shift our efforts towards ensuring that humanity survives for the long term.

1. Introduction

Although Tutankhamun has been dead for over three millennia, we have some ideas about his life. He was slight-of-build and may have walked with a cane, the unfortunate result of a curved spine. He came to the Egyptian throne at the age of nine and died about a decade later. Once thought to have been murdered, scholars now believe that his death was accidental. It was perhaps the consequence of a chariot crash (Booth 2007).

Suppose that someday we come to know much more about the life of King Tut. Suppose that Mina – some future scientist – has access to Tut’s DNA, along with information about his memories, desires, and other psychological features. And suppose that Mina creates a duplicate of Tut – Tut* – to these specifications. As this duplicate hobbles around the lab, Mina might wonder: has Tut’s life *resumed*? Or has a new life *begun*?

Some will find this question interesting. Others will not, thinking it instead *empty* or *merely verbal*. But even these others may find their interest roused by a question of a more practical flavour. Rewind, and suppose that Mina has two options. She can create Tut* who (she knows for sure) will live

* Comments and questions welcome at elliott.thornley@philosophy.ox.ac.uk

a good life, or she can create Bukayo – an entirely new person – who will live a slightly better life. Whoever she creates, other people will be unaffected. Which outcome would be better?

On one view in population axiology (and granting an assumption I discuss below), the answer is simple. The *total view* implies that it is better to create Bukayo, because that will result in greater total welfare. On *critical-set views*, the answer is not so simple. Their verdicts depend on whether Tut’s life will resume. If Tut’s life will not resume, then it is better to create Bukayo. If Tut’s life will resume, then it is better to create Tut* (on *critical-level views*) or else the two outcomes are incommensurable (on *critical-range views*).¹

In this paper, I argue that these questions of identity between lives – questions of *biographical identity* – spell trouble for critical-set views. I end with a practical implication for those aiming to promote the impartial good. If we shift our credences towards the total view, we should also shift our efforts towards reducing the risk of premature human extinction.²

2. Framework

Let a *life-episode* be an episode of a life: a stretch of a person’s life without any gaps. Your third birthday (for example) is a life-episode, as is your twentieth year, as is the next second, as is your life in its entirety. Let *biographical identity*

¹ If you think that the question ‘Will Tut’s life resume?’ is *empty* – that there is no answer to discover – then I can save you some time. Read the following argument, and then skip straight to Section 6:

- (1) On critical-set views, the relative value of outcomes depends on whether Tut’s life will resume.
- (2) ‘Will Tut’s life resume?’ is an empty question.
- (3) The relative value of outcomes cannot depend on the answer to an empty question.
- (C) Therefore, critical-set views are false.

Here’s a brief elaboration on premise (3). If a question is empty, then there is no answer to discover (not even ‘it is indeterminate’). We can at most *stipulate* answers to empty questions. But we cannot determine the relative value of outcomes by an act of stipulation.

I have some sympathy for this argument, but my case against critical-set views does not rely on it. From now on, I assume that questions of identity between lives are *substantive*: that they have answers to discover. This assumption is not intended to rule out that their answers are sometimes ‘it is indeterminate.’

² In a companion paper, I argue that questions of personal identity pose similar problems for person-affecting views. Those arguments have a similar practical upshot.

be a binary relation that obtains between two life-episodes iff they are episodes of the same life.

A life-episode's *welfare* is how good that life-episode is for the person living it. I assume that a life-episode's welfare can be represented by a real-valued function w , so that life-episode x has at least as much welfare as life-episode y iff $w(x) \geq w(y)$. I also assume that welfare is interpersonally comparable, so that we can say whether life-episode x has at least as much welfare as y even if x and y are lived by different people. And I assume that welfare is measurable on a ratio-scale, so that we can talk meaningfully about the ratios of welfare between life-episodes. Some life-episodes are good for the person living them, others are bad for the person living them, and still others are neutral for the person living them. These life-episodes are assigned positive, negative, and zero welfare scores respectively.³

A *population* is a set of lives. On the total view, we sum the welfare scores of the lives in a population to get the value of that population. A population X is at least as good as a population Y iff the value of X is at least as great as the value of Y .⁴ On critical-level views, we first subtract some positive constant from the welfare score of each life in a population and then sum the results to get the value of that population. This positive constant is the *critical level*. As with the total view, X is at least as good as Y iff X 's value is at least as great as Y 's.⁵ On critical-range views, we calculate the value of a population on a *range* of critical levels. X is at least as good as Y iff X 's value is at least as great as Y 's on every level in the critical range. If neither X nor Y is at least as good as the other, they are incommensurable, on a par, or it is indeterminate which is better.⁶ I adopt the language of incommensurability in this paper, but my discussion can be translated into other terms without significant change to its import. Following Thornley (2022), I use the term

³ In this paper, I ignore the complication that some lives may be *undistinguished* or *weakly neutral* (Gustafsson 2020; Rabinowicz 2022; Thornley 2022).

⁴ Advocates of the total view include Hudson (1987), Tännsjö (2002), and Huemer (2008).

⁵ Advocates of critical-level views include Blackorby, Bossert, and Donaldson (2005) and Bossert (2022).

⁶ Advocates of critical-range views include Broome (2004), who interprets the critical range as a range of indeterminacy, along with Qizilbash (2007; 2018) and Rabinowicz (2009), who each interpret the critical range as a range of parity.

‘critical-set views’ to refer to that class of views comprising both critical-level and critical-range views.

Here is an example to illustrate the difference between the total view, critical-level views, and critical-range views. Suppose that we can bring about either population A or population B , represented by the following sets of welfare scores:

$$A = \{5\}$$
$$B = \{2, 2, 2\}$$

On the total view, the value of A is 5 and the value of B is $2 + 2 + 2 = 6$, so B is better than A . On a critical-level view with a critical level of 4, the value of A is $(5 - 4) = 1$ and the value of B is $(2 - 4) + (2 - 4) + (2 - 4) = -6$, so A is better than B . On a critical-range view with a critical range running from 0 to 4, A and B are incommensurable, because A has greater value on a critical level of 4 and B has greater value on a critical level of 0. For concreteness, I discuss these critical-level and critical-range views below. Everything I write applies – *mutatis mutandis* – to views with critical levels and ranges occurring elsewhere. I also use the term *discount constant* to refer to the maximum amount by which a life’s welfare score is discounted in calculating the value of a population. On our example critical-level and critical-range views, the discount constant is 4.

That is all the set-up required for this paper. Onto the objections.

3. The Drop

Suppose that there exists a machine called the *LifeTransformer*. Stored on this machine is a digital file, containing all the information needed to create an entirely new person: Leah. At setting 0 on the LifeTransformer, nothing happens. Emile walks into the machine and then right back out again, entirely unchanged. At setting 1, a small cluster of cells in Emile’s brain and body are replaced with Leah’s.⁷ As a consequence, the person who walks out – call them Emile* – shares some psychological features with Leah. Perhaps Emile* has a few of Leah’s beliefs and intentions. At higher settings, larger clusters of Emile’s cells are replaced with Leah’s, and Emile* shares more psychological features

⁷ Or, rather, replaced with a small cluster of cells that would match a small cluster of cells in Leah’s brain, if Leah existed. I leave further qualifications of this kind implicit.

with Leah. At setting 1000, Emile's entire brain and body is replaced with Leah's, and Emile* is exactly like Leah in psychological respects.⁸

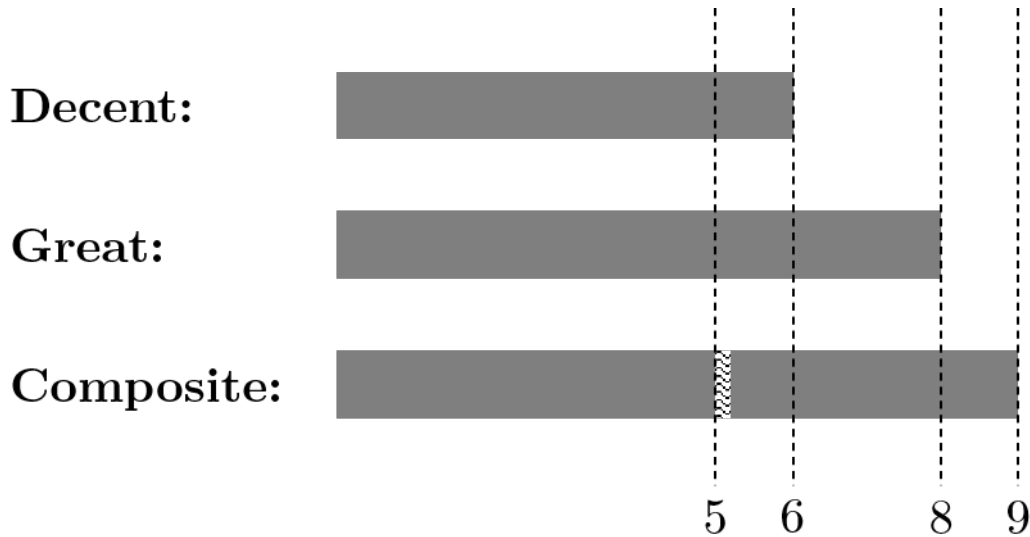
Now consider the following three outcomes:

Decent: Emile does not enter the LifeTransformer. He lives a life with a welfare score of 6.

Great: Emile does not enter the LifeTransformer. He lives a life with a welfare score of 8.

Composite: Emile lives a life-episode with a welfare score of 5. He then enters the LifeTransformer at some setting. Emile* then lives a life-episode with a welfare score of 4.

Here is a diagram to illustrate:



⁸ This case is a cosmetic variation on Parfit's *Combined Spectrum* (1984, 236–37). Like the Egyptology and Fission cases to come, it is impossible with current technology. But I claim that this impossibility does not significantly diminish the force of these cases. Neither of two common reasons for dismissing thought-experiments applies in my cases (Parfit 1984, 388). First, some claim that our moral theories need only be adequate in the actual world. Whether or not this is true, my cases do not contravene any known law-of-nature and so for all we know they could occur in the actual world. Second, some claim that some thought-experiments are impossible to imagine (see, for example, Dennett (1995, 322) on philosophical zombies). I contend that the present case and the cases to come can be imagined.

Suppose – for now – that individual welfare is *additively separable* over life-episodes: that is to say, for all non-overlapping life-episodes x and y with welfare scores $w(x)$ and $w(y)$ respectively, the life-episode composed of x and y has welfare score $w(x) + w(y)$.⁹ Consider two questions:

1. Is Composite better than Great?
2. Is Composite better than Decent?

On the total view, the answers are simple. Composite is better than Great and better than Decent. That is because (ignoring all unaffected lives), the value of Decent is 6, the value of Great is 8, and the value of Composite is 9. On our example critical-set views, the answers are not so simple. They depend on whether Emile and Emile* live the same life.

Consider first our critical-level view, with a critical level of 4. If Emile and Emile* live the same life, the value of Decent is $(6 - 4) = 2$, the value of Great is $(8 - 4) = 4$, and the value of Composite is $(9 - 4) = 5$. Therefore, if Emile and Emile* live the same life, Composite is better than Great.

If Emile and Emile* live different lives, however, the value of Decent is $(6 - 4) = 2$, the value of Great is $(8 - 4) = 4$, and the value of Composite is $(5 - 4) + (4 - 4) = 1$. The value of Composite has decreased, because we now subtract the discount constant 4 from two separate welfare scores: Emile’s and Emile*’s. Therefore, if Emile and Emile* live different lives, Composite is worse than Decent.

Clearly, when Emile enters the LifeTransformer at setting 0, he and Emile* live the same life. Equally clearly, when Emile enters the LifeTransformer at setting 1000, he and Emile* live different lives. Therefore, if biographical identity is determinately-all-or-determinately-nothing, there must be some setting k such that at k Emile and Emile* live the same life and at $k + 1$ Emile and Emile* live different lives.¹⁰ Our critical-level view then

⁹ This is the ‘assumption I discuss below’ mentioned in the introduction.

¹⁰ By ‘biographical identity is determinately-all-or-determinately-nothing,’ I mean that for every pair of life-episodes it is either determinately true or determinately false that the pair are biographically identical. The assumption thus precludes it being indeterminate whether two life-episodes are biographically identical. One might well deny this assumption: a point which I address below. That said, even given indeterminacy about biographical identity, the consequent of the footnoted conditional is still true on supervaluationism, because it is true on all precisifications. I thank an anonymous reviewer for pointing this out.

implies what-one-might-consider an implausibly large *drop* in the value of Composite as we move from k to $k + 1$. Composite goes from better than Great to worse than Decent, despite the fact that the move from k to $k + 1$ involves replacing just a few more of Emile’s cells and psychological features with Leah’s.¹¹

We get a similar drop on critical-range views. Recall that on our example critical-range view we calculate the value of each population on a range of critical levels running from 0 to 4. If Emile and Emile* live the same life, the values of Decent, Great, and Composite on a critical level of 0 are 6, 8, and 9 respectively, while their values on a critical level of 4 are 2, 4, and 5 respectively. Since the value of each population decreases linearly as the critical level increases, these values at the critical range’s endpoints imply that Composite has greater value than Great on each level in the critical range. Therefore, if Emile and Emile* live the same life, Composite is better than Great.

If Emile and Emile* live different lives, however, the values of Decent, Great, and Composite on a critical level of 0 are 6, 8, and 9 respectively, and their values on a critical level of 4 are 2, 4, and 1 respectively. The value of Composite on a critical level of 4 has decreased, because we now subtract the discount constant 4 from two separate welfare scores: Emile’s and Emile*’s. Thus, neither Composite nor Decent has at least as much value as the other on each level in the critical range. Composite has greater value on a critical level of 0 and Decent has greater value on a critical level of 4. Therefore, if Emile and Emile* live different lives, Composite is incommensurable with Decent.

If biographical identity is determinately-all-or-determinately-nothing, our critical-range view implies that there is some setting k such that Composite is better than Great (and Decent) at k and incommensurable with Decent (and Great) at $k + 1$. This change in evaluative verdicts is not as stark as the change on our critical-level view, but the drop in Composite’s value might still seem implausibly sharp. Many changes to Emile’s cells and psychological features

¹¹ This might be considered an example of what Pummer calls *hypersensitivity*: ‘when a slight difference in one sort of property makes a radical difference in another sort of property.’ (Pummer 2021, 510). In this case, it is a slight difference in microphysical and psychological properties that makes a radical difference in axiological properties.

make no difference, but one tiny change pushes Composite from better than Great to no better than Decent.

Now to consider a few objections. First, advocates of critical-set views might argue that the drop is not so bad. They might say that it should come as no surprise that Composite's value drops markedly as we move from k to $k + 1$. After all, biographical identity ceases to hold as we move from k to $k + 1$, and biographical identity matters a lot on critical-set views.¹² The point is well-taken. My claim is that once we recognise how fragile biographical identity can be (a few severed synapses and missing memories can make the difference) it becomes less plausible that it matters so much.¹³ Of course, advocates of critical-set views could *modus tollens* this *modus ponens*: since biographical identity matters a lot, a few severed synapses and missing memories can matter a lot (and indeed can matter so much that the value of Composite drops precipitously as we move from k to $k + 1$). My claim is that the *modus ponens* is more plausible.

Onto the next point. I assumed above that individual welfare is additively separable over life-episodes. That assumption allowed me to infer that, since Emile and Emile*'s welfare scores are 5 and 4 respectively when they live different lives, their combined welfare score is $5 + 4 = 9$ when they live the same life. But additive separability over life-episodes is controversial (Broome 2004, 106–9). Many philosophers believe that a life's welfare score can be greater or lesser than the sum of its parts (see, for example, Dorsey 2015 and references therein). So, it is worth noting that the drop remains a problem when we cease to assume additive separability.

Suppose first that Emile's and Emile*'s welfare score when they live the same life is greater than 9. In that case, there is still a drop. At k , Composite is better than Great. At $k + 1$, Composite is worse than Decent (on our critical-level view) or else Composite is incommensurable with Decent (on our critical-range view).

So, suppose instead that Emile's and Emile*'s welfare score when they live the same life is less than 9. In that case, so long as Emile's and Emile*'s combined welfare score is not exactly equal to 5, there will still be some

¹² I thank an anonymous reviewer for suggesting this response.

¹³ I am here still assuming that biographical identity is determinately-all-or-determinately-nothing.

discontinuity in the value of populations as we ascend the settings on the LifeTransformer. That is because, on our critical-level view, the value of Composite when Emile and Emile* live different lives is $(5 - 4) + (4 - 4) = 1$. To avoid any discontinuity whatsoever, the value of Composite when they live the same life must also equal 1. Since we subtract the discount constant 4 just once when Emile and Emile* live the same life, their combined welfare score must be 5.

More generally, to avoid all discontinuities in LifeTransformer cases on our critical-level view, the *longevity penalty* (as I will call it) must always equal 4. That is to say, whenever a life-episode y is appended to a life-episode x , the welfare score of the combined life-episode must equal $w(x) + w(y) - 4$. Then the value of a population would remain the same when life-episodes x and y came to belong to different lives, because the application of the extra discount constant would be cancelled out by the loss of the longevity penalty. But then, since even a single moment of a life is a life-episode, our critical-level view must claim that we incur a longevity penalty of 4 with each new moment. If the next moment of your life would have a welfare score of less than 4 if lived on its own, it would be better for you to die now rather than live it. That seems implausible.

Critical-range views, meanwhile, cannot avoid all discontinuities by denying additive separability. Even if Emile and Emile*'s combined welfare score is exactly equal to 5, there will still be a discontinuity. It will just be in the opposite direction: a jump rather than a drop. On a critical level of 0, the value of Composite when Emile and Emile* live the same life will be 5, while the value of Composite when they live a different life will be 9. Therefore, Composite is worse than Decent at k and incommensurable with Great at $k + 1$.

A more promising way to soften these discontinuities is to move the critical level towards 0 in the case of critical-level views, and to move one or both of the endpoints of the critical range towards 0 in the case of critical-range views. If, for example, we lower the critical level from 4 to 3, any discontinuity will be smaller. But note two points. First, the closer the critical level and critical range are to 0, the more critical-level and critical-range views behave like the total view. Second, even a small discontinuity seems implausible. The difference between Emile* at k and Emile* at $k + 1$ might be no more than a

few cells and faint memories: the kind of change that you and I undergo every minute. It is hard to believe that a population featuring Emile* at $k + 1$ is significantly worse than a population featuring Emile* at k . To avoid discontinuities entirely, we must have a single critical level at 0, and then the view renders all of the same verdicts as the total view.

A more radical way for advocates of critical-set views to avoid discontinuities is to deny another assumption that I made above. Besides assuming that individual welfare is additively separable over life-episodes, I also assumed that biographical identity is determinately-all-or-determinately-nothing: that there is some setting k on the LifeTransformer such that at k it is determinately true that Emile and Emile* live the same life and at $k + 1$ it is determinately true that they live different lives. That led me to assume that the application of the discount constant is also all-or-nothing: that at k Emile*'s welfare score is discounted by 0 and at $k + 1$ it is discounted by 4. But advocates of critical-set views can deny this last assumption. They can claim instead that the discount to Emile*'s welfare score increases in small increments as we ratchet up the settings on the LifeTransformer. Perhaps at setting 1, Emile*'s welfare score is discounted by 0.004, at setting 2, it is discounted by 0.008, and so on. That would allow critical-set views to avoid any discontinuities. As we ramp up the settings, there will come a point at which Great is better than Composite and Composite is still better than Decent on critical-level views, and a point at which Great is incommensurable with Composite and Composite is still better than Decent on critical-range views.

This *discount-by-degrees* – as I will call it – could be justified by claiming that biographical identity is sometimes indeterminate and that the truth of claims about biographical identity admits of degrees.¹⁴ A discount-by-degrees could also be justified by claiming that the size of the discount constant depends not on biographical identity but on some relation more commonly thought to come in degrees, such as psychological or physical connectedness.¹⁵ Whichever way the move is justified, however, critical-set views will face an objection from Egyptology.

¹⁴ Lewis (1976) makes these claims about personal identity.

¹⁵ Parfit (1984, 313) makes this claim of prudential decisions: the degree to which we can rationally discount future welfare depends on psychological connectedness.

4. Egyptology

The total view and critical-set views satisfy *Separability over Lives*: whether an outcome A is at least as good as an outcome B depends only on the existence and welfare of lives affected by the choice between A and B .¹⁶ Other views in population axiology – like the average view, variable value views, and rank-discounted views – do not satisfy Separability over Lives: whether A is at least as good as B can depend on the existence and welfare of lives *unaffected* by the choice.¹⁷ On these latter views, we may have to do research in Egyptology – figuring out how numerous and well-off the ancient Egyptians were – to determine which of the outcomes available to us is best. That requirement seems implausible, and many take it as a reason to reject such views.¹⁸

It is commonly thought that critical-set views – being separable over lives – do not require research in Egyptology. But that is not true. At least, it is not true so long as critical-set views are paired with what I call a *non-fire account* of biographical identity. I explain the distinction between fire and non-fire accounts below. For now, it suffices to say that, on non-fire accounts, life-episodes need not be spatiotemporally continuous to be part of the same life. Critical-set views paired with non-fire accounts require Egyptology whether they feature an all-or-nothing discount constant or a discount-by-degrees.

To see how, recall the case of Mina and Tutankhamun. Assume for now that individual welfare is additively separable over life-episodes, and suppose for concreteness that Tut’s ancient Egyptian life-episode has a welfare score of 10, Tut*’s life-episode would have a welfare score of 9, and Bukayo’s life-episode would have a welfare score of 10. On our critical-level view, creating Tut* is better than creating Bukayo iff the discount d applied to Tut*’s welfare score is less than 3, and creating Bukayo is better than creating Tut* iff d is greater than 3.¹⁹ On our critical-range view, creating Tut* is incommensurable with

¹⁶ Blackorby, Bossert and Donaldson (2005, 127) call this condition ‘Existence Independence.’

¹⁷ See Thomas (2022) and Tarsney and Thomas (2020) for discussion.

¹⁸ See McMahan (1981, 115) for the original point and Parfit (1984, 420) for the ancient Egyptians example.

¹⁹ Ignoring all unaffected lives, and supposing that Tut’s and Bukayo’s lives are entirely new and hence fully discounted, the value of creating Bukayo is $(10 - 4) + (10 - 4) = 12$, while the value of creating Tut* is $(10 - 4) + (9 - d)$. If $d < 3$, creating Tut* has more value. If $d > 3$, creating Bukayo has more value.

creating Bukayo iff d is less than 3, and creating Bukayo is better than creating Tut* iff d is greater than or equal to 3.²⁰ Therefore, on our critical-set views, which outcome is best depends on the size of the discount applied to Tut*'s welfare score. And that in turn depends on whether Tut and Tut* live the same life, or else on the extent to which Tut* resembles Tut in certain respects. Thus, on our critical-set views, Mina may need to read up on Tut's life and figure out how closely his memories, desires, and other psychological features would be matched by Tut*'s in order to determine which of the outcomes available to her is best. That requirement seems implausible.

As above, I have thus far assumed that individual welfare is additively separable over life-episodes. But, again as above, denying additive separability is an unappealing escape-route. We can avoid the need for Egyptology on critical-level views only if the longevity penalty is $4 - d$. Then in cases where Tut and Tut* are similar, the discount d is low and the longevity penalty is high, while in cases where Tut and Tut* are dissimilar, the discount d is high and the longevity penalty is low. In each case, the value of the population resulting from Mina's creating Tut* remains the same, so Mina can know the value of creating Tut* without knowing how closely Tut* resembles Tut. But, as before, this view is implausible with respect to welfare. It implies that, with each passing undiscounted moment of your life, you incur a longevity penalty of 4. If your next moment is undiscounted and would have a welfare score of less than 4 were it lived alone, it would be better for you to die now rather than live it.

Critical-range views, meanwhile, cannot avoid Egyptology by denying additive separability. If a longevity penalty cancels out the effect of a discount from some level in the critical range, it will fail to cancel out a discount from some other level. Thus, the value of Mina's creating Tut* on at least one level in the critical range – and hence whether it is better to create Tut* than some other life – will depend on how closely Tut* resembles Tut.

²⁰ Creating Bukayo is never worse than creating Tut*, because creating Bukayo has greater value on a critical level of 0: the value of creating Bukayo is $(10 - 0) + (10 - 0) = 20$ and the value of creating Tut* is $(10 - 0) + (9 - 0) = 19$. Creating Tut* is incommensurable with creating Bukayo iff $(10 - 4) + (9 - d) > (10 - 4) + (10 - 4)$, where d is the maximum discount applied to Tut*'s welfare. That is, iff $d < 3$.

Rather than avoid Egyptology, advocates of critical-set views might instead accept it. Or else they might claim that their kind of Egyptology (learning about ancient Egyptians' physical and psychological features) is acceptable, whereas the kind of Egyptology required by views like the average view (learning about how numerous or well-off the ancient Egyptians were) is unacceptable.²¹ In defending this claim, advocates of critical-set views might contend that the value-relations pertinent to Mina's choice depend on the extent to which Tut* resembles Tut. If Tut* bears little resemblance to Tut, creating Tut* is more like creating a new life and Tut*'s welfare should be heavily discounted. If Tut* bears a strong resemblance to Tut, then creating Tut* is more like bringing Tut back from the dead and Tut*'s welfare should be discounted little if at all. Bringing people back from the dead is better than creating new lives.

Even with this rationale, however, the need for any kind of Egyptology still seems to me like a blow. It seems implausible to claim that which of Mina's available outcomes is best could depend on – say – whether an ancient Egyptian Pharaoh liked the taste of honey. More implausible still is the following implication: which outcome is best could depend on the resemblance between Tut and Tut* even if Tut's life was (and Tut*'s life would be) not particularly rich or varied: even if, for example, Tut's life was (and Tut*'s life would be) no more than an unbroken period of mild and uniform pleasure.²² What is more, I expect these implications to seem especially worrying to advocates of critical-set views. After all, one of the major attractions of these views was that they seemed to avoid the need for Egyptology.²³ Of course, advocates might hold on to critical-set views even so, but these views will have lost a significant draw.

²¹ I thank an anonymous reviewer for suggesting this response.

²² This proviso rules out cases in which Tut* would complete some project of Tut's or satisfy some of Tut's desires: cases in which it might seem more plausible that the value of the available outcomes depends on the resemblance between Tut and Tut*.

²³ Wilkinson (2022, 467), for example, writes that avoiding the need for Egyptology is 'one of the main appeals' of views like critical-set views. See also remarks from Blackorby, Bossert, and Donaldson (2005, 132–33), Broome (2004, 194), and Bossert (2022, 74, 83).

What seems to me a better response is to pair critical-set views with a *fire account* of biographical identity. On fire accounts, lives are like fires.²⁴ Their identity requires both spatial and temporal continuity. Putting out a fire and then lighting another in the same place does not bring back the same fire, no matter how close the resemblance. The gap in temporal continuity means that the old fire is gone forever. Similarly for spatial continuity. A fire lit in a different place at the same instant some fire is put out is not the same fire, no matter how similar they are in other respects. On fire accounts, lives are the same. To die for an instant is to die forever.

For an example of a fire account, consider a version of McMahan's *Embodied Mind* account of personal identity (2002, chap. 1.5), amended so that it refers to lives rather than persons. On this account, biographical identity consists in the continued existence and functioning of enough of the same brain to support the capacity for consciousness.

Advocates of critical-set views can use fire accounts to address my Drop and Egyptology objections. They can *avoid* the drop by claiming that the discount applied to Emile*'s welfare score increases in small increments as we ramp up the settings on the LifeTransformer, or else they can *justify* the drop by appealing to their criterion of biographical identity. If they adopt an Embodied Mind account, for example, they can claim that the drop should come as no surprise: despite the small physical and psychological differences between Emile* at k and Emile* at $k + 1$, passing through the LifeTransformer at k preserves Emile's capacity for consciousness and passing through at $k + 1$ does not. Fire accounts also imply that Mina need not do research in Egyptology. Since there is no spatiotemporal continuity between Tut and Tut*, she can be sure that creating Tut* means creating a new life. Which of the available outcomes is best will not depend on how closely Tut* resembles Tut.

²⁴ Analogies along these lines are old. See Seneca (2004, Letter LIV, 104-5):

Wouldn't you say that anyone who took the view that a lamp was worse off when it was put out than it was before it was lit was an utter idiot? We, too, are lit and put out. We suffer somewhat in the intervening period, but at either end of it there is deep tranquillity.

See also the Aggi-Vacchagotta Sutta (*Majjhima Nikāya* 72), in which the Buddha compares asking where an enlightened person goes after death to asking where a fire goes after it is blown out.

However, trouble remains. Advocates of critical-set views may be surprised to find themselves driven towards such a narrow class of views about biographical identity. They may also be reluctant to accept some of fire accounts' implications. Consider, for example, Parfit's Teletransporter (1984, 199), which vaporises your brain and body and then creates a perfect replica out of new matter. Since the Teletransporter does not preserve spatiotemporal continuity between you and your replica, fire accounts imply that your life ends when you enter. And a variation on Parfit's Teletransporter throws up some unsavoury ethical implications. First imagine a long, wonderful life. Then suppose that some event like this life occurs, except that the brain and body at its centre are momentarily and frequently blinked out of and then back into existence. Call this event a *wonderful-but-blinking life-series*. Since fire accounts imply that each blink causes the end of one life and the beginning of another, critical-set views paired with a fire account imply that the welfare scores of each of these short lives is discounted. If the blinks occur frequently enough, the value of each wonderful-but-blinking life-series on a positive critical level will be arbitrarily low. That means that critical-level views paired with a fire account imply the *Blinking Sadistic Conclusion*:

For any population of awful lives, there is some population of wonderful-but-blinking life-series such that the blinking population is worse than the awful population.

Critical-range views, meanwhile, imply the *Weak Blinking Sadistic Conclusion*:

For any population of awful lives, there is some population of wonderful-but-blinking life-series such that the blinking population is *not better* than the awful population.²⁵

Both conclusions seem tough to accept, since the blinking population is exactly like a population of wonderful lives except for the blinks.

Advocates of critical-set views might react by holding on to fire accounts and accepting a Blinking Sadistic Conclusion, or else by rejecting fire accounts and accepting the need for Egyptology. Neither option strikes me as appealing, and both options lead to trouble in cases of fission.

²⁵ For the original Sadistic Conclusion, see Arrhenius (2000, 256). For the Weak Sadistic Conclusion, see Gustafsson (2020, 86).

5. Fission

Suppose that Asiya's brain is divided in two, and each half is implanted into an exact replica of her body. Each of the resulting people – call them Lefty and Righty – share all of Asiya's psychological features. Both Lefty and Righty are also phenomenally, physically, and functionally continuous with pre-fission Asiya. That is to say, Asiya's stream of (and capacity for) consciousness divides and flows uninterrupted into the streams of (and capacities for) consciousness of Lefty and Righty.²⁶

In this case, which – if any – of Lefty's and Righty's welfare scores is discounted? Here are six possible answers.

- (1) Both Lefty's and Righty's welfare scores are discounted.
- (2) Lefty's welfare score is discounted.
- (3) Righty's welfare score is discounted.
- (4) One of Lefty's and Righty's welfare scores is discounted, but it is indeterminate which.
- (5) Each of Lefty's and Righty's welfare scores is 'half-discounted.'
- (6) Neither Lefty's nor Righty's welfare scores is discounted.

I believe that only (6) is viable. Each of (1)-(5) implies some especially implausible Sadistic Conclusion. To see how, suppose that Asiya splits into Lefty and Righty. Each of Lefty and Righty then live a life-episode with a welfare score of 1, before themselves splitting in two. Each of their descendants

²⁶ This is a cosmetic variation on Parfit's *My Division* (1984, 254–55). Contra my claims in footnote 8, one might contend that this case – involving as it does the dividing of consciousness into two separate streams – will not happen in the actual world or is not imaginable. However, it may be interesting to note that this dividing seems to have actually occurred in patients whose corpus callosum has been severed. The evidence for this 'dual consciousness' phenomenon comes in the form of experiments like the following. Two objects are presented to a patient, one in each half of their visual field. If the patient is asked to pick out a matching object by feel alone, they pick out only the object presented in the left half of their visual field. If the patient is then asked to name the object they have picked, they name only the object presented in the right half of their visual field. As the experimenters write, 'it is as if two separate brains were viewing the left and right halves of the visual field.' (Sperry, Gazzaniga, and Bogen 1969, 278). See also Sperry's (1982) and Parfit's (1984, 245–46) discussions of this phenomenon.

also lives a life-episode with a welfare score of 1 before splitting in two, and so on. Call this a *good-but-splitting life-tree*. On answers (1)-(5) and a critical level of 4, each split reduces the population's value: each of the two splittees lives a life-episode with a welfare score of 1, but the welfare discount is at least 4.²⁷ That means that our critical-level view paired with (1)-(5) implies the *Splitting Sadistic Conclusion*:

For any population of awful lives, there is a population of good-but-splitting life-trees that is worse.

Our critical-range view paired with (1)-(5), meanwhile, implies the *Weak Splitting Sadistic Conclusion*:

For any population of awful lives, there is a population of good-but-splitting life-trees that is not better.

These Splitting Sadistic Conclusions are more troubling than the originals, since each splittee can be psychologically, phenomenally, physically, and functionally continuous with all of their ancestors and descendants. Their lives need not be lives of 'muzak and potatoes' either (Parfit 1986, 148). In fact, each splittee's life-episode can be almost exactly like an episode within a long, wonderful life. The only difference is that this life frequently branches, with each descendant also living a life-episode almost exactly like an episode within a long, wonderful life.

Thus, I take it that advocates of critical-set views will opt for (6): when Asiya splits, neither Lefty's nor Righty's welfare score is discounted. That answer allows critical-set views to avoid both forms of Splitting Sadistic Conclusion. The catch is that (6) exposes critical-set views to analogues of *all* of the problems that afflict the total view, *in addition to* the classic problems for critical-set views like the original Sadistic and Weak Sadistic Conclusions.²⁸

Consider first the Repugnant Conclusion (Parfit 1984, 388):

²⁷ On (1), each of Lefty's and Righty's welfare scores is discounted by 4, for a total discount of 8. On (2), (3), and (4), one of Lefty's and Righty's welfare scores is discounted by 4, for a total discount of 4. On (5), each of Lefty's and Righty's welfare scores is discounted by 2, for a total discount of 4.

²⁸ For other objections to critical-set views, see Thornley (2022).

For any population of wonderful lives, there is a population of lives barely worth living that is better.

The total view implies the Repugnant Conclusion, while our example critical-set views do not. However, critical-set views paired with (6) do imply the *Splitting Repugnant Conclusion*:

For any population of wonderful lives, there is a population of *life-branches* barely worth living that is better.

By ‘life-branch’ I mean the kind of life-episode lived by Lefty and Righty: life-episodes that begin post-fission. To see how critical-set views plus (6) imply the Splitting Repugnant Conclusion, consider a finite but arbitrarily large population of wonderful lives. Call this population *A*. Population *B* starts out with the same number of lives as *A*, but each life immediately splits and the welfare score of each splittee’s life-branch is half of the welfare score of the *A*-lives. *C* is similar to *B*, except that each life immediately splits twice and the welfare score of each splittee’s life-branch is a quarter of the welfare score of the *A*-lives. And so on until we reach *Z*, in which each *A*-life immediately splits many times and each splittee’s life-branch is barely worth living. Perhaps the only pleasures in each such life-branch are muzak and potatoes (Parfit 1986, 148). Z^+ is identical to *Z* but for a gumdrop’s worth of pleasure added to each life-branch. (6) states that the welfare score of each splittee’s life-branch is undiscounted. Critical-set views then imply that Z^+ is better than *A*.²⁹

²⁹ Why do I render the Splitting Repugnant Conclusion in terms of *life-branches* rather than *lives*? Because one could claim that fission preserves biographical identity: when Asiya splits into Lefty and Righty, there remains just one life (Dainton (1992) makes this kind of claim about personal identity: Asiya, Lefty, and Righty are each identical to each other). One might argue for this claim as follows: Asiya’s life-episode is biographically identical to both Lefty’s and Righty’s life-episodes, and identity is transitive, therefore Lefty’s and Righty’s life-episodes are biographically identical. One could also accept what Gustafsson and Kosonen (forthcoming) call ‘the Prudential Total View,’ on which a life’s welfare score is the sum of the welfare scores of each of its moments (even if some of those moments are lived simultaneously). These claims imply that the lives in Z^+ are *not* barely worth living. Each branch is barely worth living, but each life is wonderful in virtue of its many branches. I expect that readers will find the Splitting Repugnant Conclusion unpalatable even in light of this argument. After all, the Z^+ world could be almost exactly like the large-population world in the original Repugnant Conclusion. Both

More generally, wherever *creating new lives* presents a problem for the total view, *creating new life-branches* presents an analogous problem for critical-set views paired with (6). Consider an example. Given a plausible principle about the link between value and reasons, the total view implies that we have *reason* to create lives barely worth living. Then given a plausible principle about the link between reasons and obligations (and in the absence of any countervailing considerations), the total view implies that we are *obliged* to create lives barely worth living. That might seem implausible. However, critical-set views paired with (6) have a similarly implausible implication. Given plausible principles about the links between value, reasons, and obligations, critical-set views imply that we are obliged to create *life-branches* barely worth living.

The upshot is that fission presents a real challenge to critical-set views. If advocates claim that some discount constant applies in fission cases, critical-set views imply some Splitting Sadistic Conclusion. If, on the other hand, advocates claim that no discount constant applies in fission cases, critical-set views face analogues of all of the problems that blight the total view, in addition to the classic problems faced by critical-set views alone.

Thus, I claim, considerations of biographical identity give us reason to shift our credences away from critical-set views and towards the total view. Once we begin asking questions about identity between lives, critical-set views run into all kinds of difficulties. Paired with some claims about biographical identity, they entail implausible discontinuities in the value of populations. Paired with other claims, they require research in Egyptology to determine

could contain a vast number of human beings subsisting on muzak, potatoes, and a single gumbop. The only difference would be in origins: the human beings in Z^+ would be the product of fission. It thus seems implausible to claim that all the lives in Z^+ are wonderful, and this implausibility counts against the conjunction of the Dainton-style view about biographical identity and the Prudential Total View. What is more, one might disbelieve the conjunction on other grounds. The Dainton-style view implies that you could by chance meet someone living a life that is numerically identical to your own and that this person could clue you in on recent happenings in your own life. And given the Dainton-style view, the Prudential Total View implies an intrapersonal Repugnant Conclusion: no matter how wonderful your life, it would be even better for you to split your life into a large number of barely-worth-living life-branches in which the only pleasures are muzak and potatoes.

which outcome available to us is best. And no matter what our views about biographical identity, they have troubling consequences in fission cases.

6. Practical Implications

Suppose, then, that we shift some portion of our credence from critical-set views to the total view. This move has practical implications for those aiming to promote the impartial good.

To see how, consider an example. You have £1 billion at your disposal. As it stands, you estimate that there is a 10% chance that humanity goes extinct this century (in which case total future welfare scores will be roughly zero) and a 90% chance that 10^{16} people exist in the future, with an average welfare score of 10 in expectation.³⁰ You have two options:

1. Donate to the Nuclear Threat Initiative, and thereby reduce the risk of human extinction this century from 10% to 9.99%.³¹
2. Donate to Emergent Ventures, and thereby increase expected average future welfare scores conditional on survival from 10 to 10.01.³²

On the total view, the expected value of donating to the Nuclear Threat Initiative is $(0.099 \times 0) + (0.901 \times 10 \times 10^{16}) = 9.01 \times 10^{16}$, and the expected value of donating to Emergent Ventures is $(0.1 \times 0) + (0.9 \times 10.01 \times 10^{16}) = 9.009 \times 10^{16}$. Therefore, given expected value theory, the total view implies that it is better to donate to the Nuclear Threat Initiative.³³

³⁰ 10^{16} is Bostrom's (2013, 18) conservative estimate of future population size, conditional on avoiding near-term catastrophe.

³¹ The Nuclear Threat Initiative is a non-profit aiming to prevent global catastrophes. See www.nti.org/about for more details.

³² Emergent Ventures is a grant program aimed at funding ideas for meaningfully improving society. See www.mercatus.org/emergent-ventures for more details.

³³ Expected value theory states that an option A is at least as good as an option B iff the expected value of A is at least as great as the expected value of B , where the expected value of an option is defined as the probability-weighted average of the values of that option's possible outcomes.

There are many ways to deviate from expected value theory. For instance, theories can recommend that we instead maximise the expectation of some strictly increasing transformation of value, or they can place extra weight on some outcomes (see, e.g., Buchak 2013). These theories require that I tweak my example, but do not affect the general point that I make below.

On our critical-level view, meanwhile, the expected value of donating to the Nuclear Threat Initiative is $(0.099 \times 0) + (0.901 \times (10 - 4) \times 10^{16}) = 5.406 \times 10^{16}$, and the expected value of donating to Emergent Ventures is $(0.1 \times 0) + (0.9 \times (10.01 - 4) \times 10^{16}) = 5.409 \times 10^{16}$. So, given expected value theory, our critical-level view implies that it is better to donate to Emergent Ventures.³⁴ Since donating to Emergent Ventures has greater value on a critical level of 4 and donating to the Nuclear Threat Initiative has greater value on a critical level of 0, our critical-range view implies that the two options are incommensurable.

In this case, then, shifting some portion of our credence from critical-set views to the total view enhances the appeal of donating to the Nuclear Threat Initiative. More generally, placing more stock in the total view increases the relative importance of ensuring humanity's long-term survival and decreases the relative importance of improving humanity's prospects conditional on survival.

7. Conclusion

Critical-set views avoid the Repugnant Conclusion by subtracting some constant from the welfare score of each life in a population. These views are thus sensitive to facts about biographical identity, and this sensitivity raises a whole host of problems. If the application of the discount constant is all-or-nothing, critical-set views lead to implausible discontinuities in the value of populations. Severing one synapse and erasing one faint memory can make a population significantly worse. If biographical identity does not require spatiotemporal continuity, then critical-set views require us to become Egyptologists to determine which of some set of outcomes is best. And if biographical identity does require spatiotemporal continuity, then critical-set views imply some Blinking Sadistic Conclusion. We can add some Splitting Sadistic Conclusion to the list of charges if the welfare scores of splittees are discounted. And if the welfare scores of splittees are not discounted, critical-set

³⁴ Why the disagreement between the total view and our critical-level view? The views agree about the value of increasing welfare scores from 10 to 10.01, but they disagree about the value of adding lives at welfare 10. On the total view, the value of doing so is 10. On our critical-level view, the value of doing so is $10 - 4 = 6$.

views imply the Splitting Repugnant Conclusion instead, along with analogues of all the other problems faced by the total view.

So, I argue, we should reject critical-set views in favour of the total view. This move has practical implications for those aiming to promote the impartial good. It decreases the relative importance of improving humanity's future conditional on survival, and increases the relative importance of ensuring that humanity has a future.³⁵

8. References

- Arrhenius, Gustaf. 2000. 'An Impossibility Theorem for Welfarist Axiologies'. *Economics & Philosophy* 16 (2): 247–66.
- Blackorby, Charles, Walter Bossert, and David Donaldson. 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Cambridge: Cambridge University Press.
- Booth, Charlotte. 2007. *The Boy Behind the Mask: Meeting the Real Tutankhamun*. Oxford: Oneworld.
- Bossert, Walter. 2022. 'Anonymous Welfarism, Critical-Level Principles, and the Repugnant and Sadistic Conclusions'. In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius, Krister Bykvist, Tim Campbell, and Elizabeth Finneron-Burns. Oxford: Oxford University Press.
- Bostrom, Nick. 2013. 'Existential Risk Prevention as Global Priority'. *Global Policy* 4 (1): 15–31.
- Broome, John. 2004. *Weighing Lives*. Oxford: Oxford University Press.
- Buchak, Lara. 2013. *Risk and Rationality*. Oxford: Oxford University Press.
- Buddha, Gautama. n.d. 'Aggi-Vacchagotta Sutta'. In *Majjhima Nikāya*, 72. <https://www.dhammadata.org/suttas/MN/MN72.html>.
- Dainton, Barry F. 1992. 'Time and Division'. *Ratio* 5 (2): 102–28.
- Dennett, Daniel C. 1995. 'The Unimagined Preposterousness of Zombies'. *Journal of Consciousness Studies* 2 (4): 322–26.
- Dorsey, Dale. 2015. 'The Significance of a Life's Shape'. *Ethics* 125 (2): 303–30.

³⁵ For helpful comments and discussion, I thank Adam Bales, Hilary Greaves, William MacAskill, Teruji Thomas, two anonymous reviewers for the *Australasian Journal of Philosophy*, and the audience at the Slippery Slope Normativity Summit 2022.

- Gustafsson, Johan E. 2020. ‘Population Axiology and the Possibility of a Fourth Category of Absolute Value’. *Economics & Philosophy* 36 (1): 81–110.
- Gustafsson, Johan E., and Petra Kosonen. forthcoming. ‘Prudential Longtermism’. In *Essays on Longtermism*, edited by Jacob Barrett, Hilary Greaves, and David Thorstad. Oxford: Oxford University Press.
- Hudson, James L. 1987. ‘The Diminishing Marginal Value of Happy People’. *Philosophical Studies* 51 (1): 123–37.
- Huemer, Michael. 2008. ‘In Defence of Repugnance’. *Mind* 117 (468): 899–933.
- Lewis, David. 1976. ‘Survival and Identity’. In *The Identities of Persons*, edited by Amelie Oksenberg Rorty, 17–40. University of California Press.
- McMahan, Jeff. 1981. ‘Problems of Population Theory’. *Ethics* 92 (1): 96–127.
- McMahan, Jefferson. 2002. *The Ethics of Killing: Problems at the Margins of Life*. New York: Oxford University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- . 1986. ‘Overpopulation and the Quality of Life’. In *Applied Ethics*, edited by Peter Singer, 145–64. Oxford: Oxford University Press.
- Pummer, Theron. 2021. ‘Sorites on What Matters’. In *Ethics and Existence: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan, 498–523. Oxford: Oxford University Press.
- Qizilbash, Mozaffar. 2007. ‘The Mere Addition Paradox, Parity and Vagueness’. *Philosophy and Phenomenological Research* 75 (1): 129–51.
- . 2018. ‘On Parity and the Intuition of Neutrality’. *Economics & Philosophy* 34 (1): 87–108.
- Rabinowicz, Wlodek. 2009. ‘Broome and the Intuition of Neutrality’. *Philosophical Issues* 19 (1): 389–411.
- . 2022. ‘Getting Personal: The Intuition of Neutrality Reinterpreted’. In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius, Krister Bykvist, Tim Campbell, and Elizabeth Finneron-Burns. Oxford: Oxford University Press. <https://www.iffs.se/en/publications/working-papers/studies-on-climate-ethics-and-future-generations-vol-2/>.
- Seneca, Lucius Annaeus. 2004. *Letters from a Stoic*. Translated by Robin Campbell. London: Penguin Books.

- Sperry, Roger W. 1982. ‘Some Effects of Disconnecting the Cerebral Hemispheres’. *Science* 217 (4566): 1223–26. <https://doi.org/10.1126/science.7112125>.
- Sperry, Roger W., Michael S. Gazzaniga, and Joseph E. Bogen. 1969. ‘Interhemispheric Relationships: The Neocortical Commissures; Syndromes of Hemisphere Disconnection’. In *Handbook of Clinical Neurology*, edited by P. J. Vinken and G. W. Bruyn, 4:273–90. Amsterdam: North-Holland Publishing Co.
- Tännsjö, Torbjörn. 2002. ‘Why We Ought to Accept the Repugnant Conclusion’. *Utilitas* 14 (3): 339–59.
- Tarsney, Christian J., and Teruji Thomas. 2020. ‘Non-Additive Axiologies in Large Worlds’. <https://globalprioritiesinstitute.org/christian-tarsney-and-teruji-thomas-non-additive-axiologies-in-large-worlds/>.
- Thomas, Teruji. 2022. ‘Separability and Population Ethics’. In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius, Krister Bykvist, Tim Campbell, and Elizabeth Finneron-Burns, 271–95. Oxford: Oxford University Press.
- Thornley, Elliott. 2022. ‘Critical Levels, Critical Ranges, and Imprecise Exchange Rates in Population Axiology’. *Journal of Ethics and Social Philosophy* 22 (3): 382–414.
- Wilkinson, Hayden. 2022. ‘In Defence of Fanaticism’. *Ethics* 132 (2): 445–77.