The Neutrality of Life

ANDREW Y. LEE | University of Oslo, Philosophy

Abstract:

Some philosophers think that life is worth living not merely because of the goods and the bads within it, but also because life itself is good. I explain how this idea can be formalized by associating each version of such of a view with a function from length of life to the value generated by life itself. Then I argue that every version of the view that life itself is good faces some version of the following dilemma: either (1) good human lives are worse than very long lives wholly devoid of pleasure, desire-satisfaction, knowledge, or any other goods, or (2) very short lives containing nothing but suffering are worth living. Since neither result is plausible, we ought to reject the view that life itself is good. On the view I favor, any given life may be worth living because of the goods that it contains, but life itself is neutral.

§1   Introduction

Some philosophers think the goodness of a life is determined not only by the goods within that life, but also by goodness from life itself:

> There are elements which, if added to one's experience, make life better; there are other elements which…make life worse. But what remains when these are set aside is not merely neutral: it is emphatically positive. Therefore life is worth living even when the bad elements of experience are plentiful, and the good ones too meager to outweigh the bad ones on their own. The additional positive weight is supplied by experience itself—Nagel [1970: 2].

On this picture, even after counting all the goodness and badness from within a life, there remains some extra goodness from life itself. Imagine a scale that weighs life's goods against its bads, but where the plate weighing the goods is itself heavier: to achieve equilibrium, the weight from the bads must exceed the weight from the goods.

This paper argues that life itself is neither good nor bad. I'll explain how any theory that takes life itself to be good can be formalized by specifying a function from length of life to the value generated by life itself. I'll argue that every version of the view that life itself is good faces some version of the following dilemma: either (1) good human lives are worse than very long lives wholly devoid of pleasure, desire-satisfaction, knowledge, or any other goods, or (2) very short lives containing nothing but suffering are worth living. Since neither result is plausible, we ought to reject the view that life itself is good. On the view I favor, any given life may be worth living because of the goods that it contains, but life itself is neutral.

§2 explains what it means to say that life itself is good; **§3** develops the initial version of my core argument, which I call the 'Argument for Zero'; **§4** develops a more sophisticated version of the Argument for Zero; **§5** responds to various ways of resisting my arguments; and **§6** discusses implications for other ethical issues.

§2   The Positive Theory

For any life, we can ask how good or bad that life is, whether it's worth living, and whether it's better or worse than some other life. These questions concern what I'll call the *global value* of a life. I'll assume that a life is worth living just in case its global value is above zero, and that global value is the sum of two quantities: (1) the value due to the character of the life, and (2) the value due to life itself. In **§4**, I'll discuss views that deny the preceding claims.

The value due to the *character* of a life is the total value from the goods (and bads) that are contained within the life. By a *good*, I just mean something that makes a life intrinsically better (and by a *bad*, something that makes a life intrinsically worse). Which things are good will depend on one's preferred theory of welfare, but common candidates include pleasure, desire-satisfaction, and knowledge. Though my focus is on the value due to the entire set of goods and bads within a life (rather than any particular good or bad), I'll remain largely neutral on questions about how the values of individual

goods aggregate (such as whether the value generated by a set of goods is the sum of the values generated by those goods individually).

The value due to *life itself* is the focus of this paper.[1] The *positive theory*, which is the view I sketched at the beginning, holds that the value due to life itself is positive.[2] If the positive theory is correct, then the set of goods contained within a life is a proper subset of the set of all goods that contribute to the global value of that life (since life itself belongs to the latter, but not the former). The *neutral theory*, which is the view that I endorse, holds that the value due to life itself is zero. If the neutral theory is correct, then any particular life may be valuable or disvaluable due to its character, but life itself contributes no additional value. There is also the *negative theory*, which takes the value due to life itself to be negative, but I'll assume that such a theory is a non-starter. I'll use the phrase 'value of life' to denote the value due to life itself (rather than to denote the global value of a life), though in contexts that risk ambiguity I'll include 'itself' in the phrase.

These questions about the values of lives are fundamentally questions about welfare.[3] To have a life is to be a welfare subject, or the kind of thing that can be doing well or badly. The global value of a life is the total welfare generated by that life. The value due to the character of a life is the value due to the welfare goods and bads within that life. And if life is worth living for its own sake, then life itself is a welfare good. Speaking metaphorically, the

---

[1] Kagan [2012: 259] says: "If life per se has some positive value, then to decide how well off someone is you can't just add up the *contents* of their life…[W]e also have to add in something *extra*…the value in and of itself of being alive. So first we get the content subtotal, and then we add some extra positive points for the very fact that you are alive at all."

[2] For endorsements of the positive theory, see Brentano [1876/1973: 119], Nagel [1970], Schumacher [2010: 204], and Kriegel [2019]. For further discussions of this issue, see Kagan [2012, Ch.12] and Rantanen [2012]. For endorsements of the claim that life is intrinsically valuable, see Frankena [1973], Lamb [1988: 45], Agar [2001], Link [2013]. For some prior arguments against the positive theory, see Glover [2006] and Lee [2018].

[3] See Frankena [1973] for a list of candidate welfare goods (including life).

focus of this paper is on whether the "container" of welfare goods is itself a welfare good. For the purposes of this paper, I'll assume *welfare invariabilism*, the claim that the same theory of welfare goods is true of every welfare subject.[4] There are analogues of my arguments that target variabilist views, but assuming invariabilism will simplify the discussion.

When I talk about value, I'll always mean value that is *intrinsic* (as opposed to instrumental), *pro-tanto* (in that it's defeasible), and *personal* (in that it's for an individual). I'll treat 'goodness' as synonymous with 'positive value' and 'badness' as synonymous with 'negative value'. My arguments are neutral on most questions about the nature of value. But I'll take for granted that the values of lives can be represented by real numbers (where zero marks the threshold for a life worth living), and that for any two lives L_1 and L_2, either L_1 is better than L_2 or L_2 is better than L_2 or they are equally good. It's possible to develop versions of my arguments without these assumptions, but taking them for granted simplifies the exposition.

2.1   The Bad-Life-Worth-Living Intuition

Consider the following life:

BAD LIFE: A life of normal length whose average quality is negative.

Let the *average quality* of a life L be the value due to the character of L divided by the length of L. Average quality depends only on the value due to the character of a life (and not on any value from life itself). By contrast, let the *average value* of a life L be the global value of L divided by the length of L. If life itself is good, then average quality comes apart from average value, since life itself generates value beyond the value generated by the goods and

---

[4] See Lin [2018] for a defense of welfare invariabilism.

bads that it contains. But if life itself is neutral, then average quality and average value are equivalent.

Let the *bad-life-worth-living intuition* be the intuition that some version of BAD LIFE is worth living.[5] The explanation offered by those who endorse this intuition is that there is some goodness from life itself that offsets the net badness from within BAD LIFE. If that is correct, then even though the average quality of BAD LIFE is negative, its global value may still be positive. The additional value comes from life itself.

The bad-life-worth-living intuition has been explicitly endorsed by a number of authors, and I'll assume that it's the core motivation for the positive theory. Some, such as Kriegel [2019], claim only that a life of moderately bad quality would be worth living. Others, such as Nagel [1970], say that even a life of very bad quality would be worth living. And some, such as Schumacher [2010], suggest that just about any life, no matter how terrible its quality, would be worth living.[6] For our purposes, it won't matter how exactly we imagine the character of BAD LIFE. Let's just stipulate that the average value of BAD LIFE is *barely positive* by the lights of the positive theorist. Then the positive theorist affirms (while the neutral theorist denies) that BAD LIFE is worth living.

§3   The Argument for Zero

To develop my argument, let me first define several kinds of lives:

---

[5] Kagan [2012: 260] says: "[T]he main reason for…accepting a [positive] theory is precisely to remind us that in deciding whether you are better off dead…it may not be sufficient to focus on the *contents* of the life; it may be important to add some positive points above and beyond the content subtotal so as to take into account the value of the sheer fact that you're alive."

[6] Rasmussen & Bailey [2021] argue that persons have infinite value. Although they distinguish their thesis from the claim that the *lives* of persons have infinite value, I think a version of my arguments will apply to their view about the values of persons.

An *excellent* life: A life with an average quality very far above zero.

An *awful* life: A life with an average quality very far below zero.

An *empty* life: A life devoid of any goods or bads (except for life itself).

To imagine an excellent life, think of the best things from your own life and imagine a life filled with things that are even better. To imagine an awful life, think of the worst things from your own life and imagine a life filled with things that are even worse. To imagine an empty life, imagine a life comprised of nothing but neutral experiences of gray, with no pleasures (nor pains), no desires, no knowledge or friendship, and nothing else from a standard list of welfare goods (and bads). Now consider the following lives:

PARADISE: An excellent life of normal length.

ETERNITY: An empty life that lasts indefinitely.

SWIFT HELL: An awful life that lasts one minute.

With these lives defined, I can present my initial argument:

T   **The Basic Argument for Zero**

P1:   PARADISE is better than ETERNITY.

P2:   SWIFT HELL is not worth living.

P3:   If life itself is valuable, then either P1 or P2 is false.

—

C: Life itself is not valuable.

The argument is valid. Both P1 and P2 are highly plausible. The remaining premise is P3, which will require several stages of argument to develop and defend.

Before moving forward, let me make two preemptive remarks. First, my arguments appeal to a number of equations and functions, which I use to formalize different positive theories. These formalizations may initially strike some readers as unnecessary for a philosophical analysis. But the formalizations will eventually yield philosophical fruits: in particular, they play an essential role in my argument for why every version of the positive theory yields counterintuitive results. Second, some of my arguments appeal to infinitely long lives or infinitary values. I formulate the arguments in this way because doing so simplifies the exposition. But for those who dislike the idea of infinitary lives, it's possible to develop versions of my arguments that appeal only to finite lengths and finite values.

3.1 The Argument from Eternity

Any positive theorist must answer the following question: how does the value of life relate to the length of life? A natural answer is that more life means more value: the longer a life, the greater the value generated by life itself. This leads to the simplest and most straightforward version of the positive theory:

Linearity
The value of life increases linearly as a function of the length of life.

We can formalize *Linearity* with a simple equation relating length of life to value of life itself. To construct this equation, let me first introduce two functions—λ and τ—that will recur many times throughout this paper:

$$\lambda(L) \quad = \quad \text{the value of life itself for L}$$
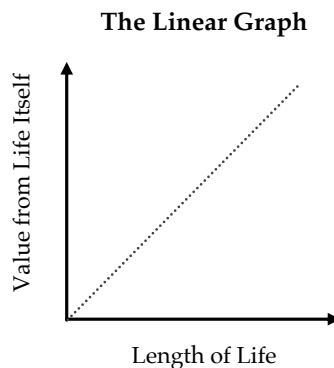$$\tau(L) \quad = \quad \text{the length of life L (in years)}$$

λ takes as input a life and outputs the value generated for it by life itself. τ takes as input a life and outputs the length of that life. The only other

element needed to specify the equation for *Linearity* is a constant *a*, which provides a way of differentiating theories that accept *Linearity* but that vary on how they scale the value of life itself relative to the length of life:

$$\lambda(\text{L}) = a \times \tau(\text{L})$$

The Linearity Equation

In the graph for the linearity equation, *a* specifies the slope of the line:

The Linear Graph

Now, *Linearity* seems like the obvious way of precisifying the positive theory. But it leads to the Argument from Eternity:

⊥ **The Argument from Eternity**

P1: ETERNITY has infinite value.

P2: PARADISE has finite value.

—

C: ETERNITY is better than PARADISE.

The reasoning is straightforward. According to *Linearity*, the value of life itself for any life L is the length of L times the constant *a*. Since any positive real number times infinity is infinity, ETERNITY is guaranteed to generate

infinite value from life itself. Since ETERNITY contains no bads, there is nothing to offset the goodness from life itself. By contrast, given that PARADISE is only finitely long, it's plausible that the global value of PARADISE is finite. So, *Linearity* leads to the conclusion that ETERNITY is better than PARADISE. But, contra that result, PARADISE is better than ETERNITY. So, we must reject one of the premises in the Argument from Eternity. In particular, we should reject P1, which entails rejecting *Linearity*.[7]

Some positive theorists may initially be tempted to embrace the result that ETERNITY is better than PARADISE. But notice that the Argument from Eternity works even when we compare ETERNITY to lives that are not only excellent, but also extremely long. Consider, for example, LONG PARADISE, whose average quality is just as excellent as that of PARADISE but which lasts 999,999,999,999,999 years. Anyone who accepts *Linearity* must likewise accept that ETERNITY is better than LONG PARADISE. But I suspect very few will be satisfied with this result.

The Argument from Eternity bears some superficial similarities to the Repugnant Conclusion, the thesis that for any world A containing a finite number of excellent lives, there is another world Z containing a greater number of lives barely worth living such that Z is better than A. However, it would be a mistake to think that the Argument from Eternity is simply a repackaged version of the Repugnant Conclusion. In response to the Repugnant Conclusion, one *cannot* reject the stipulation that each person in Z has a life barely worth living (since that is simply how the scenario is defined). By contrast, in response to the Argument from Eternity, one *can* reject the supposition that the average value of ETERNITY is positive (since that is precisely the

---

[7] Crisp [1997] invokes a structurally similar point when he asks whether hedonists can avoid the result that for any finitely long excellent life, there is an arbitrarily long life involving only small, simple pleasures that is better. Note, however, that while nearly everyone accepts that pleasure is good, it's controversial whether life itself is good. Few think that the long life of an oyster can be better than an excellent human life, and even fewer will think that ETERNITY is better than PARADISE.

premise that the neutral theorist denies). The Repugnant Conclusion generates a feeling of paradox (because it's unobvious which premise ought to be rejected); the Argument from Eternity does not (because it's obvious where the argument goes wrong). This means that the positive theorist cannot defend their view simply by appeal to the fact that there are seemingly similar puzzles in population ethics: the dialectical structures diverge.[8]

3.2 The Argument from Hell

To escape the Argument from Eternity, the positive theorist must reject *Linearity* (and more generally, any theory that takes the value of life to increase without bound as a function of length of life). Here's a natural alternative:
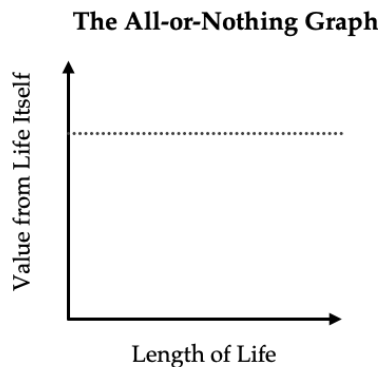
All-or-Nothing
The value of life is all-or-nothing.

According to *All-or-Nothing*, every life generates the same amount of value from life itself, regardless of its length. This view is formalized by the following equation (where *b* is the all-or-nothing value of life itself):

$$\lambda(\text{L}) = b$$

The All-or-Nothing Equation

In the graph for the all-or-nothing equation, *b* specifies the height of the line:

---

[8] These points apply to the intrapersonal version of the Repugnant Conclusion as well. Moreover, many responses to the Repugnant Conclusion (such as revising the notion of a life worth living, rejecting the transitivity of better-than, or appealing to person-affecting principles) have no obvious analogues as responses to the Argument from Eternity.

The All-or-Nothing Graph

Value from Life Itself

Length of Life

All-or-Nothing avoids the Argument from Eternity (so long as *b* is finite). But now a new issue arises: the bad-life-worth-living intuition must be restricted. Since there is no limit on how long BAD LIFE might last, there is arguably no limit on how much badness BAD LIFE might generate. This means that rejecting *Linearity* requires taking the bad-life-worth-living intuition to have a limited scope, where not all bad lives are worth living, even if we hold fixed their average quality. That result is significant, since it's natural to interpret the bad-life-worth-living intuition as a restricted universal claim, scoping over all lives whose average qualities are greater than or equal to that of BAD LIFE.

The positive theorist might counter that the bad-life-worth-living intuition is intended to hold only at the scale of human lives. After all, those who have endorsed the intuition were probably not thinking about lives of arbitrary length and character. And in any case, much of the interest of the positive theory comes from its implications for the values of our own lives. But if we restrict the bad-life-worth-living intuition, then we can ask: how much value must life itself have in order to satisfy the bad-life-worth-living intuition for normal human lives?

According to a recent report by the United Nations, the life expectancy for a person born in 2020 is about 73 years.[9] Suppose we accept both *All-or-*

---

[9] See United Nations, Dept. of Economic and Social Affairs, Population Division [2019].

Nothing and the bad-life-worth-living intuition. Then, to grasp the magnitude of the value of life, we need to appreciate the accumulation of goods and bads over the course of a whole lifetime (rather than merely the goodness or badness associated with a moment in life). Even if the average quality of BAD LIFE is only mildly negative, the all-or-nothing value of life itself would have to be extremely high in order to offset the net badness from a 73-year-long life of negative average quality.

Some might point out that even a life that is barely negative in its global value (as opposed to merely its average quality) would still count as a version of BAD LIFE. But while this point is technically correct, I think it's irrelevant in the present context. When a positive theorist thinks about BAD LIFE, they aren't thinking of a life where just one extra pinprick would tip the scale and render the life no longer worth living. Instead, the motivation for the positive theory likely comes from thinking of a moment in life that is hypothesized to be negative in quality, supposing that this level of quality reflects the average quality of the whole life, and judging that such a life would nevertheless be worth living. But then, by the reasoning above, any positive theorist who endorses *All-or-Nothing* must take life itself to have an enormous amount of value.

At first, this may appear to be good news for the positive theorist. Now they can deny that ETERNITY is better than PARADISE yet still accept that BAD LIFE is worth living. However, the escape from ETERNITY leads straight to SWIFT HELL:

SWIFT HELL: A terrible life that lasts for one minute.

How terrible is SWIFT HELL? Let's stipulate that the value due to the character of SWIFT HELL is $-b + \epsilon$, where ϵ is some very small positive number. In other words, we define SWIFT HELL to be about as awful as it can be while

still having its badness outweighed by the goodness from life itself. Speaking metaphorically, we can think of SWIFT HELL as the result of taking the net badness of BAD LIFE and condensing that into a single minute. To put that into perspective, the average quality of SWIFT HELL is a little less than 38,368,800 times as negative as the average quality of BAD LIFE.[10] Therefore, even if the average quality of BAD LIFE is only mildly negative, SWIFT HELL must be unimaginably horrendous.

Now we are ready for the Argument from Hell:

⊥   **The Argument from Hell**

P1:   The value due to life itself for SWIFT HELL is b.

P2:   The value due to the character of SWIFT HELL is $-b + \epsilon$ (where $\epsilon > 0$).

P3: b is positive.

—

C: SWIFT HELL is a life worth living.

The reasoning is straightforward. By *All-or-Nothing*, the value from life itself is b. By stipulation, the value due to the character of SWIFT HELL is $-b + \epsilon$, where ϵ is some small positive number. By the positive theory, b is positive. And by the reasoning from earlier, SWIFT HELL is unimaginably horrific. So, *All-or-Nothing* leads to the result that a very brief life containing nothing but terrible suffering is worth living. But, contra that result, SWIFT HELL is not worth living. So, we must reject one of the premises. Since P3 is the positive theory itself and P2 is a stipulation, the only option for the positive theorist is to reject P1, which is tantamount to rejecting *All-or-Nothing*.

I've now argued that *Linearity* leads to the Argument from Eternity and that *All-or-Nothing* leads to the Argument from Hell. This establishes the core of my argument against the positive theory.

---

[10] SWIFT HELL lasts one minute, BAD LIFE lasts 73 years, and there are 525,600 minutes in a year. $525{,}600 \times 73 = 38{,}368{,}800$.

§4   The Argument against Asymptote

It's possible for a positive theorist to accept both (1) that the value of life increases as a function of the length of life (as with *Linearity*) and (2) that the value of life has an upper bound (as with *All-or-Nothing*). Here's a natural way of reconciling those claims:

Asymptote
The value of life increases asymptotically with the length of life.

According to *Asymptote*, as a life grows arbitrarily long, the value generated by life itself approaches an upper bound, which I'll label c. We also need to define a new constant, n, which specifies how quickly the value of life approaches that upper bound (or, equivalently, the rate at which the marginal value of life itself shrinks). More precisely, let n mark how long a life must be in order to generate half of c:

c the maximal value that can be generated by life itself
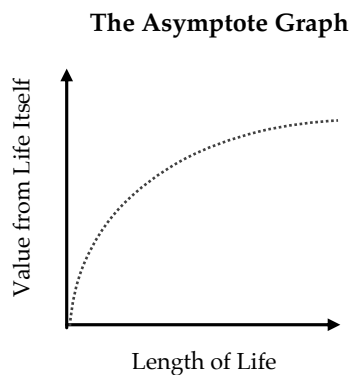n the length of life that generates half of c

Now we can specify the equation for *Asymptote*:[11]

$$\lambda(\text{L}) = \frac{c \times \tau(\text{L})}{\tau(\text{L}) + n}$$

The Asymptote Equation

In the graph for the asymptote equation, c specifies the line's vertical limit and n specifies the line's curvature:

---

[11] There are other ways of constructing an asymptotic function, but this is the simplest version that is both monotonically increasing and where $f(x) > 0$ whenever $x > 0$.

The Asymptote Graph

Here's a striking feature of the asymptote equation: the higher the value of n, the more it behaves like the linearity equation, while the lower the value of n, the more it behaves like the all-or-nothing equation. In other words, as n tends to infinity, the asymptote graph looks more and more like the linearity graph, whereas as n tends to zero, the asymptote graph looks more and more like the all-or-nothing graph. This observation will be important for my ensuing argument.

With *Asymptote*, the positive theorist can avoid both the Argument from Eternity (since there is an upper bound on the value of life itself) and the Argument from Hell (since life itself yields very little value after only a minute). Moreover, *Asymptote* satisfies the bad-life-worth-living intuition (since life itself has positive value), as well as the intuition that more life means more value (since greater lengths of life always entail greater value from life itself). Therefore, *Asymptote* initially appears to provide an escape from the Argument for Zero. But unfortunately (for the positive theorist), the appearance of escape is illusory.

Here's the basic quandary. To avoid analogues of the Argument from Eternity, one must ensure that n is not too high (or else ETERNITY will be better than a good human life). To avoid analogues of the Argument from Hell, one must ensure that n is not too low (or else SWIFT HELL will be worth living). I'll argue that the tension between these constraints cannot be resolved: every

value of n yields implausible results. These points will constitute the Argument against Asymptote.

4.1   The Setup

To set the stage for the Argument against Asymptote, we need a new function, γ, which takes as input a life and outputs the value due to the character of that life. Here again are all three of the functions we have defined:

$\lambda(\text{L})$	=	value of life itself for L
$\tau(\text{L})$	=	length of L (in years)
$\gamma(\text{L})$	=	value due to character of L

Now let's define two new lives—GOOD LIFE and SHORT HELL—which are variants of PARADISE and SWIFT HELL.[12] In addition, we will once again appeal to ETERNITY and BAD LIFE. All of these lives are characterized below (as before, let ϵ be a small positive number):

ETERNITY: An empty life that lasts indefinitely.
               $\gamma(\text{ETERNITY}) = 0$ and $\lambda(\text{ETERNITY}) = c$

BAD LIFE: A bad life that lasts for 73 years.
               $\gamma(\text{BAD LIFE}) + \lambda(\text{BAD LIFE}) = \epsilon$

GOOD LIFE: A good life that lasts for 73 years.
               $\gamma(\text{GOOD LIFE}) = -\gamma(\text{BAD LIFE})$

---

[12] It's possible to develop the Argument against Asymptote using PARADISE and SWIFT HELL, but the argument is simpler and sharper if we instead appeal to GOOD LIFE and SHORT HELL. It's likewise possible to develop the whole Argument for Zero using only GOOD LIFE and SHORT HELL, but appealing to PARADISE and SWIFT HELL better elicits the force of the Arguments from Eternity and Hell.

SHORT HELL: An awful life that lasts for 1 year.

$$\gamma(\text{SHORT HELL}) + \lambda(\text{SHORT HELL}) = \epsilon$$

_____

Here's an overview of the formal claims stated above: (1) ETERNITY is a life that contains no goods or bads but generates the maximal value from life itself. (2) BAD LIFE is a life with a negative average quality but (according to the positive theorist) a positive average value. (3) GOOD LIFE is a life where the positive value due to its character is the opposite of the negative value due to the character of BAD LIFE. (4) SHORT HELL is a stretched-out version of SWIFT HELL (or a compressed version of BAD LIFE), and is nearly as awful as it can be while still (according to the positive theorist) being worth living.

Now we are ready for the Argument against Asymptote. Recall that the asymptote equation contains a constant n, which specifies the length of life that generates half of the maximal value of life itself. I'll argue that (1) the greater the value of n, the more implausible the results become concerning ETERNITY, (2) the lower the value of n, the more implausible the results become concerning SHORT HELL, and (3) there is a middle value of n that yields implausible results for both ETERNITY and SHORT HELL. The result is that every value of n yields implausible results.

4.2   The Upward Argument

To begin, consider what happens as n becomes very large. The greater the value of n, the *smaller* the ratio between (a) the value of life itself for a life of any given finite length and (b) the maximal value of life itself. As n tends to infinity, the ratio between a and b approaches zero. Since the bad-life-worth-living intuition says that some 73-year-long bad human lives are worth living, we can think of the value of life itself at $n = 73$ as a fixed point (where the greater the value of n, the larger the maximal value of life relative to the value of life itself for a 73-year-long life). This means that the greater the value of

n, the greater the value of ETERNITY. As n tends to infinity, we approximate the Argument from Eternity. That is:

$$\lim_{n \to \infty} \lambda(\text{ETERNITY}) = \infty$$

4.3   The Downward Argument

Next, consider what happens as n becomes very small. The smaller the value of n, the *larger* the ratio between (*a*) the value of life itself for a life of any given finite length and (*b*) the maximal value of life itself. As n tends to zero, the ratio between a and b approaches 1. In other words, when n is very small, even very short lives will generate nearly the maximal value from life itself. Now, recall that $\gamma(\text{SHORT HELL}) + \lambda(\text{SHORT HELL}) = \epsilon$, meaning that the badness due to the character of SHORT HELL is barely outweighed by the value of life itself for SHORT HELL. This means that the smaller the value of n, the worse the average quality of SHORT HELL. As n tends to zero, we approximate the Argument from Hell. That is:

$$\lim_{n \to 0} \gamma(\text{SHORT HELL}) = -c + \epsilon$$

4.4   The Unstable Middle Argument

We have established that *Asymptote* encounters problems if n is either very large or very small. But is there a golden middle range that avoids implausible results in either direction? Suppose we set $n = 73$, meaning that a 73-year-long life would generate half of the maximal value of life. I'll argue that if $n = 73$, then we get implausible results concerning both ETERNITY and SHORT HELL. Then, given the Upward Argument and the Downward Argument, it will follow that no value of n avoids implausible results.

To develop this stage of the argument, it's helpful to lay out the results that occur when we apply ETERNITY, BAD LIFE, GOOD LIFE, and SHORT HELL as the arguments of the functions τ (which specifies the length of life in years),

λ (which specifies the value of life itself according to *Asymptote*, with n set to 73), and γ (which specifies the value due to the character of life). Here are those results:

Length of Life	*Value due to Life Itself*	*Value due to Character of Life*
τ(ETERNITY) = ∞	λ(ETERNITY) = c	γ(ETERNITY) = 0
τ(BAD LIFE) = 73	λ(BAD LIFE) = $\frac{1}{2}c$	γ(BAD LIFE) = $-\frac{1}{2}c + \epsilon$
τ(GOOD LIFE) = 73	λ(GOOD LIFE) = $\frac{1}{2}c$	γ(GOOD LIFE) = $\frac{1}{2}c - \epsilon$
τ(SHORT HELL) = 1	λ(SHORT HELL) = $\frac{1}{74}c$	γ(SHORT HELL) = $-\frac{1}{74}c + \epsilon$

These results, I'll argue, are unacceptable.[13]

First, consider ETERNITY and GOOD LIFE. Recall that global value is the sum of the value due to the character of life and the value of life itself: formally, *global value*(L) = λ(L) + γ(L). The results above show that if $n = 73$, then the global value of ETERNITY is slightly greater than the global value of GOOD LIFE. In particular, the global value of ETERNITY is c whereas the global value of GOOD LIFE is $c - \epsilon$. However, it's implausible that ETERNITY is better than GOOD LIFE. If you could choose which life to live, GOOD LIFE would be vastly preferable to ETERNITY. Therefore, if $n = 73$, we have an implausible result concerning ETERNITY.

Second, consider SHORT HELL and BAD LIFE. Recall that average quality is the value due to the character of life divided by length of life: formally, *average quality*(L) = γ(L) ÷ τ(L). The results above show that if $n = 73$, then the average quality of SHORT HELL is nearly twice as negative as the average quality of BAD LIFE. In particular, if we factor out the negligible value associated

---

[13] Most of these results are straightforward to verify. To see that λ(SHORT HELL) = $\frac{1}{74}c$, note that λ(SHORT HELL) = $\frac{c \times \tau(\text{SHORT HELL})}{\tau(\text{SHORT HELL}) + n} = \frac{c \times 1}{1 + 73} = \frac{c}{74} = \frac{1}{74}c$. To see that γ(SHORT HELL) = $\frac{1}{74}c + \epsilon$, recall that γ(SHORT HELL) + λ(SHORT HELL) = ϵ. Since λ(SHORT HELL) = $\frac{1}{74}c$, we can solve for γ(SHORT HELL), yielding the result that γ(SHORT HELL) = $\frac{1}{74}c + \epsilon$.

with the ϵ term, then the average quality of SHORT HELL is $-\frac{1}{74}c$ whereas the average quality of BAD LIFE is $-\frac{1}{146}c$. Yet the positive theorist is forced to say that SHORT HELL is worth living, since we stipulated that the badness due to the character of SHORT HELL is outweighed by the value of life itself for SHORT HELL. However, it's implausible that a very short life whose average quality is much worse than that of BAD LIFE is worth living. Therefore, if $n = 73$, we have an implausible result concerning SHORT HELL.

To summarize: Anyone who accepts *Asymptote* must specify a value for n, which marks how long a life must be in order to generate half the maximal value from life itself. The Upward Argument showed that if we increase the value of n, then the predictions about ETERNITY grow more implausible. The Downward Argument showed that if we decrease the value of n, then the predictions about SHORT HELL grow more implausible. The Unstable Middle Argument showed that if $n = 73$, then we get the implausible results that (1) ETERNITY is better than GOOD LIFE, and (2) SHORT HELL is worth living. Therefore, every value of n yields implausible results. So, we ought to reject *Asymptote*.

4.5 The Argument for Zero

Here's a natural question: if the positive theorist already takes BAD LIFE to be worth living, then why couldn't they likewise think that SHORT HELL would also be worth living? Well, remember that we defined BAD LIFE to be as bad as it could possibly be while still being worth living by the lights of the positive theorist. Yet whatever BAD LIFE looks like, SHORT HELL is approximately twice as bad with respect to average quality and much shorter with respect to length. This means that if the positive theorist accepts *Asymptote*, then for whatever kind of life they initially think has an average quality that renders it barely worth living, there are *shorter* lives with *worse* average qualities that are also worth living. Nearly everyone will find this result unappealing.

In fact, for similar reasons, *Asymptote* generates the result that some lives with *better* average qualities and *longer* lengths than BAD LIFE are *not*

worth living. Let LONG BAD LIFE be a life lasting 100,000 years with an average quality slightly better than the average quality of BAD LIFE (though still negative). Since LONG BAD LIFE is much longer than BAD LIFE, the badness from the character of the former will be many times greater than the badness from the character of the latter. But given *Asymptote* (with $n = 73$), the value from life itself for LONG BAD LIFE will be less than twice the value from life itself for BAD LIFE. This means that *Asymptote* entails that LONG BAD LIFE isn't worth living, despite the fact that it has a better quality and longer length than BAD LIFE. I suspect few positive theorists will be satisfied with this result.

There are, of course, infinitely many other functions from length of life to value of life. But I cannot think of any other functions that are credible candidates for constructing a positive theory.[14] And given my arguments, it's reasonable to think that every version of the positive theory will be vulnerable to some version of the Argument for Zero. The source of the problem lies not with the particular functions we have considered, but instead with the positive theory itself.

Here's the upshot: *Linearity* leads to the Argument from Eternity, *All-or-Nothing* leads to the Argument from Hell, *Asymptote* leads to a version of either (or both), and other versions of the positive theory seem susceptible to the same dilemma. Now I can state a generalized version of the Argument for Zero:

---

[14] Some might wish to appeal to the idea that the function from length of life to value of life is a sigmoid curve, where (*a*) very short lives generate little value from life, (*b*) normal-length lives generate a lot of value from life, and (*c*) extremely long lives generate only slightly more value from life than normal-length lives. For limits of space, I won't be able to address this view. But in brief, such a view leads to either (1) the conclusion that LONG BAD LIFE is not worth living, or (2) a version of the original dilemma, concerning ETERNITY and SHORT HELL.

T   **The Generalized Argument for Zero**

P1:   GOOD LIFE is better than ETERNITY.

P2:   SHORT HELL is not worth living.

P3:   If life itself is valuable, then either P1 or P2 is false.

—

C: Life itself is not valuable.

§5   Arguments against Alternatives

Let's now turn to some other options for resisting the Argument for Zero.

5.1   Option 1: Positivity

I've assumed that the only factor relevant to determining the value of life itself is quantity of life. But what if the positive theorist instead appeals to some other factor? Some may be tempted to appeal to a principle like *Diversity*, according to which the value of life increases as a function of its diversity of experiences. On this view, the value of life depends in part on how rich and variegated a life is. ETERNITY contains only a single kind of experience and SHORT HELL contains nothing but suffering. These lives lack the texture and flavor that characterize normal human lives. By contrast, even though BAD LIFE has a negative average quality, it might still be diverse enough to yield plenty of value from life itself.

The problem is that *Diversity* leads to new kinds of counterintuitive results. Let DIVERSE HELL be a life filled with an extremely diverse set of experiences, all of which are bad. Though DIVERSE HELL contains no good experiences, it still contains a rich variety of experiences: pain, hunger, thirst, anger, sadness, anxiety, frustration, disgust, horror, nausea, itchiness, and so on. DIVERSE HELL involves suffering in as many ways as you can imagine and in many more ways you cannot imagine. If *Diversity* were true, then DIVERSE HELL may well be worth living, so long as the intensity of each bad experience were sufficiently mild. But DIVERSE HELL is obviously not worth living, so we should reject *Diversity*.

This pattern of argument generalizes. If the value of life is taken to be a function of some quantity besides length of life, then we can always generate analogues of the Argument for Zero by considering lives that score high on that quantity yet are filled with bads. The only way to avoid that result would be to take the value of life to increase only as a function of the good (rather than the bad), as with the following view:

Positivity
The value of life increases as a function of its quantity of goods.

Positivity predicts that both DIVERSE HELL and SHORT HELL are not worth living (since they contain no goods), that the value of ETERNITY is zero (since it contains no goods), and yet that BAD LIFE may nevertheless be worth living (since it may still contain many goods). Since *Positivity* entails that an increase in the bads within a life will always yield a net decrease in the global value of that life, the principle is immune to the kind of argument that all the other principles have been vulnerable to.

 Nevertheless, I think there is a deeper problem with *Positivity*. We began this paper with the view that life itself is good, meaning that any life acquires some goodness from life itself, regardless of its character. But *Positivity* says instead that some lives generate no goodness at all, and that the goodness from life itself is determined by (rather than independent from) its character. That is quite a departure from the view we started with. If x is good, then it's natural to think that the amount of goodness from x must be a function of the amount of x. But *Positivity* says that even though it's x that is good, it's only more of y that generates more goodness from x. As an analogy, imagine someone who claims that belief is valuable, but who says that the value of belief itself is defeated whenever a belief is false, unjustified, unsafe, or otherwise does not amount to knowledge. That seems suspicious: if beliefs generate value when and only when they amount to knowledge, then why attribute the value to belief (rather than to knowledge)? Similarly, if

quantity of life has no direct correlation with the amount of value generated by life itself, then why attribute the value to life itself, rather than to whatever serves at the scaling factor?

5.2   Option 2: Inferiority

Let's say good g_1 is *inferior* to good g_2 (or bad b_1) just in case any amount of g_1 is worse than (or outweighed by) any amount of g_2 (or b_1).[15] And let's say g_1 is *comparable* to g_2 (or b_1) just in case neither is inferior to the other. Now consider the following view:

> **Inferiority**
> Life itself is inferior to every good and bad within life.

If *Inferiority* is true, then GOOD LIFE is better than ETERNITY (since ETERNITY's only good is life itself, which is inferior to the goods within GOOD LIFE) and neither SHORT HELL nor DIVERSE HELL is worth living (since life itself is inferior to the bads within those lives). However, this apparent solution brings about a new problem: *Inferiority* cannot retain the bad-life-worth-living intuition. If life itself is inferior to any bad within life, then no version of BAD LIFE is worth living. In fact, *Inferiority* predicts that even a single pinprick can outweigh any value accrued from life itself. Only *balanced lives*, where the goods (within the life) exactly counterweigh the bads, lead to divergent predictions between *Inferiority* and the neutral theory: *Inferiority* takes balanced lives to be barely above (rather than exactly on) the threshold of being worth living. For any other life, the value due to life itself may as well be zero. This strikes me as tantamount to giving up on the positive theory.

What if the positive theorist contends that life itself is inferior to only some (rather than all) goods and bads within life? This weakening would enable the positive theorist to retain a form of the bad-life-worth-living

---

[15] See Arrhenius & Rabinowicz [2015] for a recent discussion of inferiority principles.

intuition, but it no longer renders the positive theory immune to the Argument for Zero. If some goods are comparable to life itself, then we could consider a version of GOOD LIFE containing only those kinds of goods in order to generate the result that ETERNITY is better than GOOD LIFE. Or if some bads are comparable to life itself, then we could consider a version of SHORT HELL containing only those kinds of bads in order to generate the result that SHORT HELL is worth living. As soon as the positive theorist weakens *Inferiority*, the Argument for Zero returns.

At this point, some may be tempted to simply contend that the value due to life itself is incommensurable with the value due to the character of life. Perhaps there are two distinct dimensions of value and no objective fact of the matter about how to compare the two dimensions. However, while rejecting commensurability nullifies the Argument for Zero, it also nullifies the bad-life-worth-living intuition. In order to accept incommensurability, one must give up the intuition that motivates the positive theory in the first place. Moreover, the positive theorist cannot retreat to the idea that life is worth living so long as one's life has a positive value along at least one of the two dimensions, for that would mean that lives such as SWIFT HELL (as well as lives that are much worse) are worth living.

Finally, the positive theorist might point out that the results from the Argument for Zero are less counterintuitive if we take the value of life itself to be extremely small. But softening the blow of the counterintuitive results likewise saps the interest from the positive theory itself. We began this paper with a philosophically provocative picture, where even a life where the bads are plentiful and the goods are scarce would be worth living because of the extra goodness from life itself. That picture gradually dissolves as the value of life gradually fades to nothing.

5.3   Option 3: Anti-Globalism

My arguments have assumed the following principle:

Globalism:

Life L_1 is better than life L_2 just in case *global value*(L_1) > *global value*(L_2).

I defined global value as the sum of (1) the value due to the character of a life, and (2) the value due to life itself. But one might contend that there are other facts that determine whether one life is better or worse than another. Consider, for example, the idea that a life that gets better over time is better (all else equal) than a life that gets worse over time, even if both lives contain exactly the same set of goods and bads. At first, it may seem that these sorts of factors are excluded by *Globalism*. However, while the paradigms of the goods within life are atomistic goods like pleasure, desire-satisfaction, and knowledge, none of my remarks preclude holistic factors such as shape of life from contributing to the value due to the character of a life. The only stipulation I've made is that the value due to the character of life excludes any value from life itself. This means my arguments are compatible with a variety of views about which factors make lives better or worse.

A second approach is to appeal to average values instead of global values. Recall that the average value of a life is the global value of that life divided by the length of that life. According to *Average,* life L_1 is better than life L_2 just in case *average value*(L_1) > *average value*(L_2). Now, *Average* is a principle about what makes one life better than another (rather than a principle about the value of life itself), but we can combine it with a principle like *Linearity* to get a version of the positive theory. In fact, *Average + Linearity* yields promising results: the average value of BAD LIFE is barely positive, the average value of ETERNITY is lower than the average value of GOOD LIFE, and the average value of SHORT HELL is negative. However, consider SWIFT HEAVEN, a life that lasts for one minute but that has an average value slightly better than the average value for PARADISE. *Average + Linearity* predicts that SWIFT HEAVEN is better than PARADISE. But that is the wrong result: PARADISE is better than SWIFT HEAVEN. Therefore, we ought to reject *Average + Linearity*. As far as I

can see, other principles about the value of life fare no better when combined with *Average*.[16]

A third approach is to appeal to the idea that life L₁ is better than life L₂ just in case the *scaled value* (rather than the global value or average value) of L₁ is greater than that of L₁. The notion of scaled value is technically complex and has been systematically discussed in Arrhenius [2000],[17] so I'll keep my discussion brief. We can think of scaled value as average value times *scaled length*, where scaled length approximates length of life for short lives but approaches a maximal bound for longer lives. This makes scaled value behave like global value when comparing lives of very short length and like average value when comparing lives of very long length.[18] As a result, an appeal to scaled value can secure the results that GOOD LIFE is better than ETERNITY, SHORT HELL is not worth living, BAD LIFE is worth living, and SWIFT HEAVEN is worse than PARADISE.[19] The problem, though, is that scaled value generates new counterintuitive results. These include the results that (1) for any life L, there is some positive average value α and some negative average value β such that it would be better (all else equal) for L to be extended by some small number of years with (negative) average value β rather than

---

[16] The challenges for average welfare principles are familiar from the population ethics literature. See, as examples, Parfit [1984: 420] and Huemer [2008]. Note also that even those sympathetic to average utilitarianism (about populations of lives) tend to still endorse *Globalism* (about individual lives). See, for example, Pressman [2015].

[17] See also Ng [1989] and Sider [1991].

[18] To formally define scaled value, we must first specify a constant k between 0 and 1, where higher values of k make scaled value closer to global value and lower values of k make scaled value closer to average value. We then use k to define a new function ω from lives to *scaled lengths*, where $\omega(\text{L}) = \sum_{i=1}^{\tau(\text{L})} k^{i-1} = k^0 + k^1 + k^2 + \cdots + k^{\tau(\text{L})-1}$. Then *scaled value*(L) = *average value*(L) × ω(L). Note that this formula is structurally identical to Ng [1989]'s Variable Value Principle: the only difference is that Ng's population variable has been replaced in the formula above with a length of life variable.

[19] Actually, I suspect even positive theories that appeal to scaled value will face versions of the Argument for Zero. But I'll set aside this worry.

some large number of years with (positive) average value α, (2) lives that are worth living can sometimes be made worse just by adding years with positive average value, and (3) lives that are not worth living can sometimes be made better just by adding years with negative average value. These results make me think that an appeal to scaled value cannot save the positive theory.

5.4   Out of Options?

I've argued against a variety of approaches for resisting the Argument for Zero, and I cannot think of any other credible ways of developing the positive theory. Given this, I believe that the Argument for Zero is sound and that the positive theory is false. Supposing that the negative theory is a non-starter, the only viable option is the neutral theory. This concludes my argument for the neutrality of life.

§6   Ethical Implications

What implications does the neutral theory have for other issues in ethics?

A central task in ethics is to develop the correct theory of welfare. This theory ought to tell us, amongst other things, which things make one better or worse off. By arguing for the neutral theory, I've argued that the "container" of welfare goods is not itself a welfare good. A more subtle consequence concerns how we quantify the values of other welfare goods. I've argued that no version of BAD LIFE is worth living. But perhaps the life that the positive theorist imagines when they endorse the bad-life-worth-living intuition isn't actually a life with a negative average quality. If that turns out to be the case, then the positive theorist need not revise their views about which particular lives are worth living—instead, they need only revise their views about how to quantify the values of those lives.

If we move from individual lives to populations of lives, then we encounter the sorts of issues addressed in population ethics. An interesting observation is that if the positive theory were true, then the Repugnant Conclusion would be "super-repugnant," since the two theses in conjunction would

mean that for any finite population A containing only excellent lives, there is a larger population Z containing only lives with *negative* average qualities such that Z is better than A. The neutral theory avoids super-repugnance because it entails that whether a life is worth living depends only on the goods and bads within that life, meaning that only lives with positive average qualities are worth living.

The neutral theory also has implications for a number of issues within applied ethics. The idea that life itself is valuable is sometimes appealed to as justification for claims about the wrongness of euthanasia, abortion, meat-eating, murder, or suicide, or for claims about the goodness of procreation or life extension. The neutral theory doesn't settle whether such claims are true or false, but it does constrain the options for arguing for such claims. Whatever one thinks about those issues, one ought not appeal to the value of life itself as justification for one's view.

It's worth highlighting a special connection to the question of whether consciousness is intrinsically valuable.[20] To have a life, in the sense at stake here, is to be a welfare subject. This definition leaves open which things have lives, meaning what it is for something to be a welfare subject. But a common view is that consciousness is what makes an entity a welfare subject, meaning that all and only conscious entities have lives. Given such a view, the Argument for Zero likewise establishes the neutrality of consciousness. This paper may thus be reinterpreted as arguing that consciousness has zero intrinsic value (or disvalue).

Note that my arguments about the value of life are independent from the issues investigated in the literatures on (*a*) the value of existence, and (*b*) the badness of death. The former concern whether some outcomes can be better or worse for one than not existing at all. The latter concern whether death can be bad for the one who dies, given that one no longer exists after death. These literatures focus on issues like the *non-identity problem*, the

---

[20] See Lee [2018].

puzzle of whether bringing an individual into existence with a life barely worth living would be wrong if the individual would not exist otherwise, and on views like the *deprivation theory*, according to which death is bad because of the goods that it deprives one of. While the neutral theory has implications for what counts as a life barely worth living and which goods death deprives one of, it leaves open which actions are permissible in non-identity scenarios and whether the deprivation theory is true in the first place.[21]

Although I've argued that life is neutral, my arguments remain compatible with views that take life to be ethically significant in other kinds of ways. Consider, as examples, the ideas that (1) anything that has a life thereby has moral status, or (2) there is a right to life. One might think that life confers moral status not because life itself is good, but instead because having a life makes one a welfare subject and because all welfare subjects have moral status. Or one might think that one has a right to life not because life itself is good, but instead because life is what enables one to acquire goods in the first place.

§7   Conclusion

This paper has shown how different versions of the view that life itself is valuable can be associated with different functions from length of life to value of life. By formalizing the idea, we can utilize the formal structure to expose implausible consequences of the relevant views. I've argued that the positive theory leads to a dilemma: either (**1**) good human lives are worse than very long lives wholly devoid of pleasure, desire-satisfaction, knowledge, or any other goods, or (**2**) very short lives containing nothing but

---

[21] There is also the question of whether existence itself is good, which *is* analogous to the question of whether life itself is good. My point is simply that the philosophical literatures labeled 'value of existence' and 'badness of death' tend not to focus on the kinds of issues addressed in this paper.

suffering are worth living. I suspect most who were initially attracted to the bad-life-worth-living intuition will be reluctant to embrace either option. Those who still favor the positive theory must choose a horn of the dilemma.

On the picture I favor, life itself is neither good nor bad. To determine how good a life is, or whether a life is worth living, or whether one life is better than another, we need only look at the goods and bads within that life. In other words, the goodness of a life is determined only by the goods and bads within a life, for life itself is neutral.[†]

ANDREW Y. LEE
University of Oslo, Dept. of Philosophy, Classics, History of Ideas & Art

References

Agar, N., 2001. *Life's Intrinsic Value*, New York: Columbia University Press.

Arrhenius, G., 2000. An Impossibility Theorem for Welfarist Axiology, *Economics and Philosophy*, 16: 247–266.

Arrhenius, Gustaf and Rabinowicz, Wlodek 2015. Value Superiority. In Iwao Hirose & Jonas Olson (eds.), *The Oxford Handbook of Value Theory*. New York, USA: Oxford University Press USA. 225-248.

Brentano, F.C. 1876. *The Foundation and Construction of Ethics*. Trans. E.H. Schneewind. Routledge & Kegan Paul, 1952/1973.

Crisp, Roger. 1997. *Routledge Philosophy Guidebook to Mill on Utilitarianism*. Routledge.

Frankena, William K., 1973, *Ethics*, second edition, Englewood Cliffs: Prentice Hall.

Glover, Jonathan. 2006. The sanctity of life. In Helga Kuhse & Peter Singer (eds.), *Bioethics: An Anthology*. Blackwell. 266--275.

Huemer, Michael. 2008. In defence of repugnance. *Mind* 117 (468):899-933.

Kagan, Shelly. 2012. *Death*. Yale University Press.

Kriegel, Uriah. 2019. The Value of Consciousness. *Analysis* 79 (3):503-520.

Lamb, David. 1988. *Down the Slippery Slope: Arguing in Applied Ethics*. Routledge.

Lee, Andrew Y. 2018. Is consciousness intrinsically valuable? *Philosophical Studies* 175 (1):1–17.

Lin, Eden. 2018. Welfare Invariabilism. *Ethics* 128 (2):320-345.

Link, Hans-Jürgen. 2013. Playing God and the Intrinsic Value of Life: Moral Problems for Synthetic Biology? *Science and Engineering Ethics* 19 (2):435-448.

Nagel, Thomas. 1970. Death. *Noûs* 4 (1):73-80.

Ng, Y.-K.. 1989. "What Should We Do About Future Generations? Impossibility of Parfit's Theory X", *Economics and Philosophy*, 5: 135–253.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford University Press.

Pressman, Michael. 2015. A Defence of Average Utilitarianism. *Utilitas* 27 (4):389-424.

Rantanen, Rosa. 2012. Considerable Life Extension and the Deprivation View. *Thanatos*. 1 (2).

Rasmussen, Joshua and Bailey, Andrew M. 2021. How Valuable Could a Person Be? *Philosophy and Phenomenological Research* 103 (2):264-277.

Schumacher, Bernard N. 2010. *Death and Mortality in Contemporary Philosophy*. Cambridge University Press.

Sider, Theodore R. 1991. Might theory X be a theory of diminishing marginal value? *Analysis*, 51: 265—71.

United Nations, Department of Economic and Social Affairs, Population Division 2019. World Population Prospects 2019: Highlights. ST/ESA/SER.A/423.