—————————

The Neutrality of Life

—————————

Andrew Y. Lee | andrewyuanlee@gmail.com | http://andrewyuanlee.com/
University of Oslo, Department of Philosophy

ABSTRACT:

There are some who 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 this 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 of normal length 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. I conclude by discussing the implications for other issues in normative and applied ethics.

Introduction

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

> 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 1979: 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 the goods in life against the 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. Whether a life is worth living depends not only on which goods and bads that life contains, but also on the additional value from life itself.

This paper aims to (1) show how the question of whether life itself is good can be assessed rigorously and systematically, and (2) argue that life itself is neither good nor bad. I'll explain how different versions of the view that life itself is good correspond to different functions from length of life to value of life itself. Then I'll argue that every version of the view that life itself is good faces some version of the following dilemma: either good human lives of normal length are worse than very long lives wholly devoid of pleasure, desire-satisfaction, knowledge, or any other goods, or very short lives containing nothing but suffering are worth living. Since these results are implausible, 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.

The question of whether life itself is valuable has connections to issues about the threshold for a life worth living, the basic welfare goods, and analogous debates about the value of consciousness and the moral significance of existence. The question also connects to a variety of issues in the applied ethics of life and death, including issues concerning euthanasia, abortion, suicide, procreation, vegetarianism,

murder, life extension, and the harm of death. These connections will be discussed briefly towards the end of the paper.

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

§1 | The Positive Theory

Let's begin by setting up the basic conceptual framework for this paper. For any life, we can ask how good or bad that life is, whether that life is worth living, and whether that life is better or worse than some other life. These questions concern what I'll call the *global value* of a life. For most of this paper, 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.[1]

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.[2] The *positive theory*, which is the view I sketched at the beginning, holds that the value of life itself is positive.[3]

---

[1] In §4, I discuss views that appeal to factors besides (1) and (2).

[2] 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."

[3] For endorsements of the positive theory, see Brentano [1876/1973: 119], Nagel [1979], Schumacher [2010: 204], and Kriegel [2019]. For further discussions of this issue, see Kagan

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 of 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, of course, also the *negative theory*, which takes the value of life itself to be negative, but I'll assume that such a theory is a non-starter. In general, I'll use the phrase 'value of life' to denote the value of life itself (rather than to denote the global value of a life), though in contexts that risk ambiguity I include 'itself' in the phrase.

These questions about the values of lives are fundamentally questions about welfare.[4] What it is 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 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 the truth of *welfare invariabilism*, the view that the same theory of welfare goods is true of every welfare subject.[5] It's possible to generate 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). In general, I'll use '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

---

[2012, Ch.12] and Rantanen [2012]. For endorsements of the claim that life is intrinsically valuable, see Frankena [1973], Lamb [1998: 45], Agar [2001], Link [2013]. For some prior arguments against the positive theory, see Glover [1977] and Lee [2018].

[4] See Campbell [2016] and Crisp [2017] for overviews of issues about well-being. See Frankena [1973] for a list of candidates for welfare goods (where life is an item in the list).

[5] See Lin [2018] for some persuasive arguments in favor of welfare invariabilism.

living), and that for any two lives A and B, either A is better than B or B is better than A 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.

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 defined as the value due to the character of L divided by the length of L. Average quality is sensitive to only the value due to the character of a life (and not the value from life itself). By contrast, let the *average value* of L be defined as the global value of L divided by the length of L. If life itself is good, then average quality can come apart from average value, since life itself generates value beyond that generated by the goods and bads within life. But if life itself is neutral, then average quality and average value are equivalent.

The *bad-life-worth-living intuition* is the intuition that some version of BAD LIFE is worth living.[6] 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 bad-life-worth-living intuition has been explicitly endorsed by a number of philosophers, and I'll take for granted 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 [1979], imply that even a life of very bad quality would be worth living. And some,

---

[6] Kagan [2012: 260] says: "[S]ince we are adding extra positive points for the fact that you're alive, then even if the *contents* subtotal is negative, the *grand* total could still be positive…Indeed, the 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."

such as Schumacher [2010], seem to suggest that just about any life, no matter how terrible its quality, would be worth living.[7]

For the purposes of my arguments, it won't actually matter how we imagine the character of BAD LIFE. But let's stipulate that the average value of BAD LIFE is *just barely positive* by the lights of the positive theorist. In other words, the average badness of BAD LIFE is defined to be just barely outweighed by its average goodness (after we count the goodness from life itself). Then the positive theorist affirms (while the neutral theorist denies) that BAD LIFE is worth living.

Some readers may question the appeal of the positive theory, perhaps because they themselves lack the bad-life-worth-living intuition. From my own experience, however, a good number of philosophers are drawn to the bad-life-worth-living intuition, including but not limited to the authors mentioned above who have explicitly endorsed the positive theory. Others may wonder whether anything could be said to persuade the positive theorist, since the bad-life-worth-living intuition may seem like a basic intuition rather than a thesis reached through philosophical analysis. However, I suspect that many who find themselves initially tempted by the positive theory will reconsider once its consequences become clear. Moreover, anyone who still favors the positive theory in the end will be forced to accept some implications that will strike most as unpalatable.

§2 | The Argument for Zero

The Argument for Zero appeals to the following kinds of lives:

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 much better. To imagine an awful life, think of the worst things from your own life and imagine a life filled with things that are

---

[7] Relatedly, Rasmussen & Bailey [forthcoming] argue that persons have infinite value. Although they distinguish their thesis from the claim that the *lives* of persons have infinite value, I think my arguments apply also to their view about the values of persons.

much worse. To imagine an empty life, imagine a life comprised of nothing but neutral experiences of gray, with no pleasures (nor pains), no desires (satisfied or frustrated), no knowledge or friendship, and nothing else from a standard list of welfare goods (or 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 the initial version of my argument:

T   **The 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 extremely plausible, and I'll simply take them for granted. The remaining premise is P3, which will require several stages of argument to adequately 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 some of the exposition. But for those who dislike appeals to the infinitary, it's possible to develop versions of my arguments that appeal only to finitely long lives and finite values.

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 suggests the simplest and most straightforward version of the positive theory:

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

If *Linear* is true, then it's easy to specify an equation that relates the value of life to the length of life. But first, this is a good point to introduce two functions—λ and τ—that will recur many times throughout this paper:

λ(L) = the value of life itself for L
τ(L) = the length of life L (in years)

λ takes as input a life and outputs the value generated by life itself for that life. τ takes as input a life and outputs the length of that life. The only other element needed to specify the equation for *Linear* is a constant a, which provides a way of differentiating theories that accept *Linear* but vary on how they scale the value of life itself relative to the length of life. With these elements defined, we can characterize *Linear* with a simple equation:

$$\lambda(L) = a \times \tau(L)$$

The Linear Equation

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

The Linear Graph
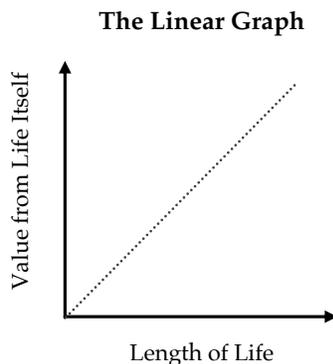
Now, *Linear* may at first seem like a natural way of precisifying the positive theory. But it leads to the Argument from Eternity:

⊥   **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 *Linear*, 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. Therefore, the global value of ETERNITY is infinite. By contrast, given that PARADISE is only finitely long, it's highly plausible that the global value of PARADISE is finite. So, if we accept P1 and P2, we are driven to the conclusion that ETERNITY is better than PARADISE. But that is the wrong result: PARADISE is obviously better than ETERNITY. Therefore, we must reject one of the premises in the Argument from Eternity: in particular, we should reject P1.

The Argument from Eternity has 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 (or more generally, the Argument for Zero) is simply a

repackaged version of the Repugnant Conclusion. In the case of 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 the case of the Argument from Eternity, one *can* reject the supposition that the average value of ETERNITY is positive (since that is precisely the 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]

The Argument from Hell

To escape the Argument from Eternity, the positive theorist must reject *Linear* (and more generally, any theory that takes the value of life itself to increase without bound). A natural alternative is the following view:

*Constant:* The value of life is a constant.

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

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

The Constant Equation

In the graph of the constant equation, c 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, and appealing to person-affecting principles) have no obvious analogues as responses to the Argument from Eternity. For more on the Repugnant Conclusion, see Arrhenius, Ryberg, & Tännsjö [2017].

The Constant Graph
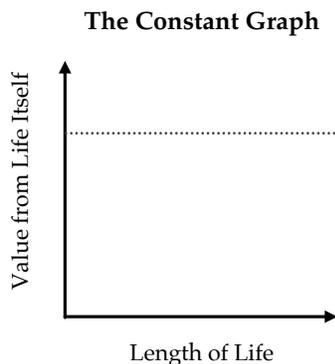
With *Constant*, we avoid the result that ETERNITY has infinite value (so long as *c* is finite). But we immediately run into a new issue: the bad-life-worth-living intuition must be restricted. Since there is no limit on how long BAD LIFE might last, there is arguably no limit on how much badness BAD LIFE might generate.[9] This means that rejecting *Linear* 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.

A natural response is for the positive theorist to say 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. So, let's suppose that 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 human lives?

---

[9] At least, this result holds if we grant the plausible assumption that adding bads to a life can increase the badness of that life without bound. Recall that the average quality of BAD LIFE is just barely negative. Let γ be a function from a life L to the value due to the character of L, and let *c* be the maximal value of life itself. Since BAD LIFE can be arbitrarily long (and thus be arbitrarily bad), it's guaranteed that for some version of BAD LIFE, $\gamma(\text{BAD LIFE}) + c < 0$.

According to a recent report by the United Nations, the global life expectancy for a person born in 2020 is about 73 years.[10] If the value of life is a constant (and if bad human lives are worth living), then grasping the magnitude of that constant requires us to think about the accumulation of goods and bads over the course of a whole lifetime (rather than merely the goodness or badness associated with a momentary experience). This means that even if the average quality of BAD LIFE is only mildly negative, the value of life itself must be remarkably high in order to offset the balance of badness from a life of negative average quality lasting 73 years.

There may be a temptation to argue that even a life that is just barely negative in its global quality would still count as a version of BAD LIFE. Because of this, the (constant) value of life itself need not be particularly high to offset the badness of BAD LIFE. But while this point is technically correct, I think it's mostly irrelevant in the present context. When a positive theorist thinks about BAD LIFE, I doubt they are thinking of a life where (say) just one extra pinprick would tip the scale and render the life no longer worth living. Instead, I suspect the motivation for the positive theory usually comes from thinking of a moment in life that is hypothesized to be negative in quality, imagining that that level of quality is reflective of the life as a whole, and judging that such a life would nevertheless be worth living. But if that is right, then by the reasoning above, any positive theorist who endorses *Constant* should be inclined to think that the value of life itself is very high.

At first, this all appears to be good news for the positive theorist. Now the positive theorist can deny that ETERNITY is better than PARADISE yet accept that BAD LIFE is worth living. But the exit from the Argument from Eternity leads straight to the Argument from Hell. Consider the following life:

––––––––––

SWIFT HELL: An awful life that lasts for one minute.

––––––––––

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

––––––––––––––––––––––––––––––

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

outweighed by the goodness from life itself. Speaking somewhat metaphorically, we can think of SWIFT HELL as the result of taking the net badness from 73 years of BAD LIFE and condensing that into a single minute (and then tweaking it down just a bit). To put that into perspective, the average quality of SWIFT HELL is a little bit less than 38,368,800 times as negative as the average quality of BAD LIFE.[11] This suggests that even if the average quality of BAD LIFE is only mildly negative, SWIFT HELL must be unimaginably horrific. And now comes the Argument from Hell:

⊥   **The Argument from Hell**

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

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

P3: c is positive.

—

C: SWIFT HELL is a life worth living.

The reasoning is straightforward. By *Constant*, the value from life itself for SWIFT HELL is the constant c. By the positive theory, c is positive. By stipulation, the value due to the character of SWIFT HELL is $-c + \epsilon$, where ϵ is some small positive number. Since any life with a global value greater than zero is a life worth living, it follows that SWIFT HELL is worth living. Now, I argued above that SWIFT HELL is an unimaginably horrific life. Therefore, this version of the positive theory must accept that a very brief life containing nothing but horrific suffering is worth living. That is the wrong result: SWIFT HELL is obviously not worth living.

To resist the Argument from Hell, the positive theorist must reject one of its premises. P3 is the positive theory itself and P2 is a stipulation. Therefore, the only option for the positive theorist is to reject P1, which is tantamount to rejecting *Constant*. The upshot is that both *Linear* and *Constant* lead to unacceptable results: the former to the Argument from Eternity and the latter to the Argument from Hell. This establishes the core of my argument against the positive theory. The aim of the rest of the paper is to generalize this dilemma.

---

[11] Where does this number come from? SWIFT HELL lasts one minute, BAD LIFE lasts 73 years, and there are 525,600 minutes in a year. $525,600 \times 73 = 38,368,800$.

§3 | 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 *Linear*) and (2) that the value of life has a maximal bound (as with *Constant*). The most natural way of reconciling these claims is with the following view:

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

As before, let *c* denote the maximal value of life itself. According to *Asymptote*, as a life grows arbitrarily long, the value generated by life itself approaches *c*. We also need to define a new constant, *n*, which will specify how quickly the value of life approaches its maximal 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 of life itself

*n*   the length of life that generates half of *c*

As an example, suppose we set *c* = 100 (meaning the maximal bound for the value of life itself is 100) and *n* = 73 (meaning that it takes 73 years for a life to generate half of the maximal value from life itself). Then the value generated by life itself converges to 100 as life gets arbitrarily long, with 73 years marking the length of time needed to generate 50 value from life itself. With both *c* and *n* defined, we can specify the equation for *Asymptote*:[12]

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

The Asymptote Equation

---

[12] There are other ways of constructing an asymptotic function, but this is the simplest version that is both constantly increasing and where $f(x) > 0$ whenever $x > 0$. Though I won't directly consider other asymptotic functions, the kind of argument I make in this section will generalize to other functions, and I'll later make some remarks that apply to all functions from length of life to value of life.

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

The Asymptote Graph
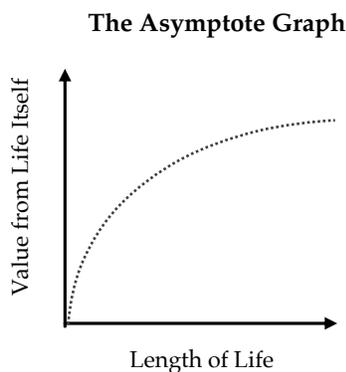
Here's a striking feature of the asymptote equation: the higher the value of n, the more it behaves like the linear equation, while the lower the value of n, the more it behaves like the constant equation. In other words, as n tends to infinity, the asymptote graph looks more and more like the linear graph, whereas as n tends to zero, the asymptote graph looks more and more like the constant graph. This observation will be important for the ensuing argument.

Now, with *Asymptote*, the positive theorist can avoid both the Argument from Eternity (since there is a maximal bound for 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). At first, it appears that *Asymptote* provides 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 an analogue of the Argument from Eternity, one must ensure that n is not too high (in order to avoid the result that ETERNITY is better than a good human life). To avoid an analogue of the Argument from Hell, one must ensure that n is not too low (in order to avoid the result that SWIFT HELL is worth living). I'll argue that the tension between these constraints cannot be resolved: every value of n yields implausible results.

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:

λ(L) = value of life itself for L
τ(L) = length of L (in years)
γ(L) = value due to character of L

Next, let's define two new lives—GOOD LIFE and SHORT HELL—which may be thought of as variants of PARADISE and SWIFT HELL.[13] 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.
                 γ(ETERNITY) = 0 and λ(ETERNITY) = c

BAD LIFE: A bad life that lasts for 73 years.
                 γ(BAD LIFE) + λ(BAD LIFE) = ϵ

GOOD LIFE: A good life that lasts for 73 years.
                 γ(GOOD LIFE) = − γ(BAD LIFE)

SHORT HELL: A bad life that lasts for 1 year.
                 γ(SHORT HELL) + λ(SHORT HELL) = ϵ

_____

_____

[13] 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.

Here's a brief description of the lives defined above: (1) ETERNITY is a life that contains no goods or bads but generates the maximal value from life itself. The formal description says that the value due to the character of ETERNITY is zero while the value of life itself for ETERNITY is maximal. (2) BAD LIFE is a life that is barely worth living by the lights of the positive theorist. The formal description says that the net negative value due to the character of BAD LIFE is barely outweighed by the positive value from life itself for BAD LIFE. (3) GOOD LIFE is a generalized version of PARADISE. The formal description says that the positive value due to the character of GOOD LIFE is the inverse of the negative value due to the character of BAD LIFE.[14] (4) SHORT HELL is a stretched-out version of SWIFT HELL (and a compressed version of BAD LIFE), and is defined as being nearly as awful as it can be while still being worth living. The formal description says that the negative value due to the character of SHORT HELL is barely outweighed by the positive value due to life itself for SHORT HELL.

Now we are ready for the Argument against Asymptote. As a reminder, the asymptote equation contains a constant n, which specifies the length of life that generates half of the maximal value from life itself. I'll argue that every value of n yields implausible results. My argument proceeds in three stages. First, I argue that the greater the value of n, the more implausible the results become concerning ETERNITY. Second, I argue that the lower the value of n, the more implausible the results become concerning SHORT HELL. Third, I argue that there is a middle value of n that yields implausible results for both ETERNITY and SHORT HELL.

The Upward Argument

To begin, consider what happens when n is 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

---

[14] Though GOOD LIFE and BAD LIFE are symmetrical with respect to the value due to their characters, the positive theory entails that they are asymmetrical with respect to global value: BAD LIFE is barely worth living (because the goodness of life itself barely outweighs the net badness due to the character of the life) while GOOD LIFE is very much worth living (because the goodness of life itself just adds further goodness on top of the net goodness due to the character of the life).

between a and b approaches zero. Since the bad-life-worth-living intuition says that a bad human life of 73 years is 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 life of 73 years). 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.

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

As n tends to infinity, the value of ETERNITY increases unboundedly

The Downward Argument

Next, consider what happens when n is 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 γ(SHORT HELL) + λ(SHORT HELL) = ϵ, 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. And as n tends to zero, we approximate a version of the Argument from Hell.

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

As n tends to zero, the quality of SHORT HELL decreases

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 life lasting 73 years would generate half of the maximal value that can be generated from life itself. 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 for ETERNITY, BAD LIFE, GOOD LIFE, and SHORT HELL when we apply to those lives 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*
$\tau(\text{ETERNITY}) = \infty$	$\lambda(\text{ETERNITY}) = c$	$\gamma(\text{ETERNITY}) = 0$
$\tau(\text{BAD LIFE}) = 73$	$\lambda(\text{BAD LIFE}) = \frac{1}{2}c$	$\gamma(\text{BAD LIFE}) = -\frac{1}{2}c + \epsilon$
$\tau(\text{GOOD LIFE}) = 73$	$\lambda(\text{GOOD LIFE}) = \frac{1}{2}c$	$\gamma(\text{GOOD LIFE}) = \frac{1}{2}c - \epsilon$
$\tau(\text{SHORT HELL}) = 1$	$\lambda(\text{SHORT HELL}) = \frac{1}{74}c$	$\gamma(\text{SHORT HELL}) = -\frac{1}{74}c + \epsilon$

These results, I'll argue, are unacceptable.

Consider first 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 were to have a choice about which life to live, it's obvious that GOOD LIFE would be vastly preferable to ETERNITY. Therefore, if $n = 73$, we have an implausible result concerning ETERNITY.

Consider next SHORT HELL and BAD LIFE. Recall that average quality is the value due to the character of life divided by the 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, 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.

It may be tempting to think that if the positive theorist already takes BAD LIFE to be worth living, then they may likewise also think that a life with an average quality that is twice as bad as BAD LIFE would also be worth living. But remember that we defined BAD LIFE as being 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. I think nearly everyone will find this result unappealing.

Therefore, $n = 73$ yields implausible results both with respect to ETERNITY and with respect to SHORT HELL.

The Argument Against Asymptote

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 of the maximal value from life itself. 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, a very short life whose average quality is nearly twice as negative as that of BAD LIFE, is worth living. The Upward Argument showed that if we raise the value of n, then the results become worse with respect to ETERNITY. The Downward Argument showed that if we lower the value of n, then the results become worse with respect to SHORT HELL. Therefore, every value of n yields implausible results. So, we ought to reject *Asymptote*.

There are, of course, infinitely many other functions from the length of life to the value of life. But I cannot think of any functions besides *Linear, Constant*, and *Asymptote* that strike me as credible candidates for a positive theory. In fact, there is reason to think that any function from length of life to value of life will be susceptible to some form of the Argument for Zero. For any such function, we can ask which lengths of life are *optimal*, meaning which lengths generate the maximal value from life itself. Then we can either (1) compare an empty life (containing no goods or bads) of optimal length to a good life of non-optimal length (as we did with ETERNITY and GOOD LIFE) or (2) consider a life of optimal length that has negative average quality but barely positive average value (as we did with SHORT HELL). No matter

what the function looks like, it seems at least one of these procedures will generate implausible results. In light of this, I suspect that every version of the positive theory is vulnerable to 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: *Linear* leads to the Argument from Eternity, *Constant* leads to the Argument from Hell, *Asymptote* leads to some version of either (or both), and other versions of the positive theory seem susceptible to the same dilemmas. This leads to a generalized version of the Argument for Zero:

⊤   **The 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.

§4 | Arguments against Alternatives

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

Option 1: Positivity

My arguments have assumed that the only factor relevant to determining the value of life itself is quantity of life. But what if the positive theorist instead takes the value of life to be determined by some other factor? It may at first be tempting 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 itself depends not on the quantity of life, but instead 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, frustra-tion, 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 is sufficiently mild. But DIVERSE HELL is obviously not worth living, so we should reject *Diversity*.

In fact, it's easy to see how 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 consider a life that scores high on that quantity yet is filled with bads. So long as that is possible, we can generate analogues of the Argument for Zero. 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), like with the following view:

Positivity: The value of life itself increases as a function of the amount of goods within it.

Positivity predicts that both DIVERSE HELL and SHORT HELL are not worth liv-ing (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). In fact, I think there is no simple counterexample to *Positivity*: the principle is immune to the kind of argument that all of the other principles have been vulnerable to. This is because *Positivity* is designed so that an increase in the bads within a life will always yield a net decrease in the global value of that life. However, I think there is a deeper problem with principles like *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 the specific character of that life. But *Positivity* says instead that some lives generate no goodness at all (even though life itself is good) and that the value of life itself is determined by (rather than independent of) the specific character of life. That is quite a departure from Nagel's idea that life is worth living even when the goods are "too meager to out-weigh" the bads because there is "additional positive weight" from life itself. If we consider again the metaphor of the scale whose plates weigh the good against the bad, *Positivity* gives us a picture where the two plates are perfectly balanced before we add any goods or bads to either side, but where the plate weighing the goods

itself becomes heavier after goods are added on top of it. That picture is awkward, and its axiological structure is inelegant.

As an analogy, imagine a philosopher who claims that belief (rather than knowledge) is valuable while also claiming that the value of belief is defeated whenever a belief is false, unjustified, unsafe, or otherwise does not amount to knowledge. It would be natural to ask: if beliefs generate value when and only when they amount to knowledge, then why not think that it's simply knowledge that is valuable? Unless there is independent reason for thinking that the value is due to belief itself, we should instead favor the simpler hypothesis that the value is due to knowledge. Similarly, any theory that contends that the value of life itself is a function of a factor besides quantity of life faces a burden of explanation: why think that the value is due to life itself, rather than due to whichever quantity serves at the scaling factor? Unless we have good reason for believing a theory of this kind, we should think that the only candidates for positive theories are those that take the value of life to be a function of the quantity of life. If we grant that quantity of life is equivalent to length of life, then we arrive back at the Argument for Zero.

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). 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.[15]

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). But now the problem is that *Inferiority* must reject the bad-life-worth-living intuition. If life itself is inferior to any bad within life, then no version

---

[15] See Arrhenius & Rabinowicz [2015] and Parfit [2016] for some recent papers that appeal to inferiority principles in order to resist the Repugnant Conclusion. Note that making that move in response to the Repugnant Conclusion is compatible with my arguments here.

of BAD LIFE is worth living. In fact, if *Inferiority* is true, then there is only one respect in which the predictions of the positive theory differ from the predictions of the neutral theory: namely, the positive theory takes empty lives (which contain no goods nor bads) to be barely above the threshold of being worth living while the neutral theory takes empty lives to be exactly on the threshold of being worth living. In any other case, the value of life itself might as well be zero. Once again, this is a very different kind of view than the one introduced at the beginning of this paper.

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

Option 3: Anti-Globalism

My arguments have assumed the following principle:

Globalism: Life A is better than life B just in case *global value*(A) > *global value*(B).

I've taken for granted that global value is the sum of (1) the value due to the character of a life, and (2) the value due to life itself. But one might challenge this assumption by contending that global value is also determined by other factors. 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, one might think that these kinds of factors are excluded by my characterization of global value. However, nothing I've said precludes factors like shape of life from contributing to the value due to the character of a life. Though the paradigms of the goods within life are atomistic goods like pleasure, desire-satisfaction, and knowledge, my arguments are compatible with the existence of holistic goods such as shape of 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 A is better than life B just in case *average value*(A) > *average value*(B). This 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 *Average* with a principle like *Linear* to get a version of the positive theory. In fact, *Average + Linear* is promising: the average value of ETERNITY is negligible, the average value of GOOD LIFE is positive, the average value of BAD LIFE is barely positive, and the average value of SHORT HELL is no worse than the average value of BAD LIFE (though note that the former is still much worse in average quality than the latter). Therefore, *Average + Linear* provides a way out of the dilemma set forth by the Argument for Zero. However, now the dilemma turns into a trilemma. Let SWIFT HEAVEN be a life that lasts for one minute and that has an average value slightly higher than the average value of PARADISE. Then *Average + Linear* predicts that SWIFT HEAVEN is better than PARADISE. But PARADISE is obviously is better than SWIFT

HEAVEN. Therefore, we ought to reject *Average + Linear*. And as far as I can see, other principles about the value of life itself fare no better when combined with *Average*.[16]

A third approach is to appeal to the idea that life A is better than life B just in case the *scaled value* (rather than the global value or average value) of A is greater than that of B. The notion of scaled value is technically complex and has already been examined systematically in Arrhenius [2000], 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.[17] As a result, an appeal to scaled value can secure the results that GOOD LIFE is better than ETERNITY, that SHORT HELL is only slightly worse than BAD LIFE, that BAD LIFE is worth living, and that SWIFT HEAVEN is worse than PARADISE.[18]

The problem is that scaled value leads to new counterintuitive consequences. 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 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 kinds of results

---

[16] The challenges for average welfare principles are familiar from prior discussions in population ethics, such as Parfit [1984: 420] and Huemer [2008]. For a recent defense of average utilitarianism (though, notably, not *Average*), see Pressman [2015].

[17] 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 $\omega(L) = \sum_{i=1}^{\tau(L)} k^{i-1} = k^0 + k^1 + k^2 + \cdots + k^{\tau(L)-1}$. Then *scaled value*(L) = *average value*(L) × ω(L). Note that this formula is structurally identical to the one used in 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.

[18] Actually, I suspect any positive theory appealing to scaled value will face a dilemma analogous to the one developed in the Argument against Asymptote. But I'll set aside this worry.

are argued for persuasively in Arrhenius [2000], and lead me to think that no appeal to scaled value can save the positive theory.[19]

Are there other ways of rejecting *Globalism*? I don't know of any other options that are promising. In fact, it's worth noting that *Globalism* is nearly always taken for granted in the philosophical literature, even by those sympathetic to principles such as average utilitarianism.[20]

Option 4: No Bad Lives Worth Living

Could the positive theorist simply forfeit the bad-life-worth-living intuition? The problem with this move is that the Argument for Zero leaves open how to specify the details of bad life. Even if we allow the value of life itself to be arbitrarily small, we could still consider versions of bad life whose average qualities or lengths are arbitrarily close to zero. Then some versions of bad life will still be worth living, and the positive theorist once again faces the Argument for Zero. Now, the positive theorist could point out that the results are less counterintuitive if we were to take the value of life itself to be extremely small. However, softening the blow of the counterintuitive results carries the cost of sapping 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 itself gradually fades to nothing.

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 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.

---

[19] Alongside Arrhenius [2000]'s criticisms of scaling principles, see Ng [1989] and Sider [1991] for some examples of appeals to scaling principles in population ethics.

[20] As an example, Pressman [2015] defends average utilitarianism (which concerns populations of lives) yet still endorses *Globalism* (which concerns individual lives).

§5 | Conclusions

I'll conclude by briefly discussing how the neutral theory connects to some other issues in normative and applied 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, eliminating one of the candidates for the basic welfare goods.[21] 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 run into 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-re-pugnant," since the two theses in conjunction would mean that for any finite population A containing only lives with excellent average qualities, 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. More generally, our judgments about which situations are better or worse in population ethics often depend upon our prior grasp of the welfare of the individuals within those situations. By constraining the factors that determine the global values of lives, the neutral theory likewise constrains the factors that determine the global values of populations.

---

[21] See, for example, Kagan [2012, Ch. 12]'s discussion of whether life itself is a welfare good, or Frankena [1973]'s list of candidates for welfare goods. The question of whether life itself is good is also closely related to recent discussions of whether consciousness is intrinsically valuable (Glannon [2016], Lee [2018], Kriegel [2019]), especially given the common idea that what it is to have a life (in the sense at stake here) is to be conscious (Nagel [1970], Sumner [1996: 43], Kahane & Savulescu [2009], Rosati [2009]).

Within applied ethics, the neutral theory has implications for a cluster of issues concerning life and death, including issues about euthanasia, abortion, suicide, procreation, vegetarianism, murder, life extension, and the harm of death. In brief, the neutral theory suggests that euthanasia is not wrong merely in virtue of the fact that it involves the ending of a person's life, that appeals to the value of life itself cannot justify the claim that abortion is wrong, that loss of life is not itself a reason against suicide, that creation of life is not itself a reason for procreation, that the justification for vegetarianism ought to focus on the harms conferred to animals rather than loss of life itself, that the badness of murder is not grounded merely in the fact that one deprives the victim of life, that life extension is not valuable for its own sake, and that death need not involve the loss of any goods. On this last point, the neutrality of life suggests an analogous neutrality of death.[22]

It's worth noting that the neutral theory crosscuts debates that fall under the label "the value of existence." Those debates focus on the question of whether some outcomes can be better or worse for one than not existing at all. The philosophical issues within those debates usually stem from puzzles such as the *non-identity problem*, the question of whether bringing an individual into existence with a low but still positive quality of life would be wrong if the individual would not exist otherwise. Since these kinds of puzzles about existence arise whether one accepts the positive theory or the neutral theory, debates about the value of existence are largely independent from questions about the value of life itself.[23]

On a similar note, a neutral theorist might still think that life is ethically significant in other ways. Consider, as examples, the idea that (1) anything that has a life thereby has moral status, or the idea that (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

---

[22] See Luper [2019] for a recent overview of the philosophy of death. See Rantanen [2012] for discussion of connections between the value of life itself and the harm of death.

[23] See Arrhenius & Rabinowicz [2015] for an overview on these puzzles about existence. Of course, 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 literature labeled 'value of existence' tends not to focus on the kinds of issues addressed in this paper.

but instead because life is what enables one to acquire goods, just as one might think that one has a right to freedom not because freedom itself is good but instead because freedom is what enables one to acquire goods.

This paper began by describing a scale that weighs the goods within life against the bads within life. If life itself is good, then the plate weighing the goods is itself heavier than the plate weighing the bads. The methodology of this paper has been to formalize different versions of this idea and to utilize that formal structure to expose implausible consequences of the view. I've argued that the positive theory leads to a dilemma: either (1) good human lives of normal length are worse than very long empty lives wholly devoid of any goods, or (2) very short lives whose average qualities are significantly worse than that of a bad human life are worth living. I suspect most who were initially attracted to the bad-life-worth-living intuition would be reluctant to embrace either of these options. Anyone who remains committed to 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.

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

Arrhenius, Gustaf, Ryberg, Jesper and Tännsjö, Torbjörn, "The Repugnant Conclusion", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/>.

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

Glover, Jonathan (1977). Causing Death and Saving Lives.

Griffin, James (1986). Well-Being: Its Meaning, Measurement and Moral Importance. Clarendon Press.

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

Jaworska, Agnieszka and Tannenbaum, Julie, "The Grounds of Moral Status", The Stanford Encyclopedia of Philosophy (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/grounds-moral-status/>.

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.

Nagel, Thomas (2012). Mortal Questions. Cambridge University Press.

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.

Parfit, Derek (2016). Can We Avoid the Repugnant Conclusion? *Theoria* 82 (2):110-127.

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

Rasmussen, Joshua & Bailey, Andrew M. (forthcoming). How valuable could a person be? Philosophy and Phenomenological Research.

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.