

# How to Live a Life Worth Living

Campbell Brown

February 26, 2008

[This is a draft; please don't cite without author's permission.]

## Abstract

Although ubiquitous in population ethics, the notion of a “life worth living” resists easy analysis. Intuitively, one wants to say that a life is worth living just in case living it is better than living no life at all. On reflection, though, this seems mysterious. To live no life at all is simply not to exist, to be nothing. But then it seems we have an instance of the “better than” relation in which one of the relata is absent; we're trying to compare something, a life, with nothing. This paper proposes an analysis of lives worth living that avoids such mysterious comparisons.

## 1. Introduction: the Puzzle of the Missing Relatum

If life is a gift, then sometimes it's a gift one should not be pleased to have received. Some lives are so undesirable as to be, in the common phrase, “not worth living”. What makes these lives especially undesirable is not merely that they're worse, or even that they're *much* worse, than other lives; a life much worse than another may nonetheless be worth living. It is, rather, that living one of these lives is no better, perhaps worse, than living no life at all.

So much, it seems, is the common view among philosophers. This idea of a life worth living, as a life “better than nothing”, is commonly employed in population ethics. As I want to suggest, however, it is a puzzling idea. To live no life at all is simply not to exist, to be nothing. Thus when we say of a life that it is better to live this life than to live no life at all, we appear to be comparing something, a life, with nothing. Somehow we have an instance of the “better than” relation in which one relatum has gone missing. This seems incoherent.

My aim here is to propose a solution to the puzzle, an analysis of lives worth living which requires no mysterious comparisons between something and nothing. After setting out my proposal, I shall compare it with an alternative analysis suggested by John Broome, and show that, given two assumptions, the analyses are equivalent. Finally, I shall consider some implications of my discussion of lives worth living for certain substantive theories of well-being.

## 2. A Proposed Solution

A good way to solve any puzzle is to begin by solving an easier version of the puzzle, and then to think of some way to “extend” the solution to cover also the more difficult version. This is the strategy I shall pursue here. I shall begin by giving an analysis of what it is for a

*future* of a life to be worth living, and then I shall show how the analysis is naturally extended from futures to whole lives.

## 2.1. Implicit Baselines and Futures Worth Living

We may make some progress in understanding futures worth living by thinking first of other contexts in which the phrase “better than nothing” occurs. We often say things that, superficially, seem to imply a comparison between something and nothing. You hoped to paint the whole fence today, but only got half done. “Oh well,” you say, “at least that’s better than nothing.” But here the comparison is really between one thing and another, not between something and nothing. There is an implicit *baseline* situation, and the thing said to be “better than nothing” is a certain alteration or addition to that baseline. In effect, the comparison is between the altered situation and the baseline. In the example, the baseline is the unpainted fence, and what you mean by your statement—“at least that’s better than nothing”—is that adding to the baseline, by painting half the fence, is better than leaving it as it is. Or, in other words, the fence half-painted is better than the fence not painted at all. Very roughly, then, “ $x$  is better than nothing” means that the baseline-plus- $x$  is better than the baseline alone.

An “implicit baseline” analysis of this sort is well suited to explaining what it is for smaller parts of lives, in particular, *futures* of lives, to be worth living. A woman loses her legs in a car accident. This is a tragedy, to be sure; nonetheless, we might think, she was lucky to have escaped death. Although her future holds great adversity, it’s still better than nothing. Here the implicit baseline is that part of the woman’s life which she has lived already, up to and including the accident. We judge that adding to this baseline by attaching to it a future of adversity is better than leaving it as it is, with no future attached. The whole life, including the future, is judged better than the life that would remain were the future removed.

Thus the notion of a *future* worth living is comparatively unproblematic. However, from the fact that a life contains a future worth living we cannot infer that life itself is worth living. Suppose your life is nothing but torture except for the last five minutes, which is bliss. When you get to the last five minutes, it’s worth living, but it’s not worth living through the rest of the life to get there in the first place. Your life as a whole is not worth living, though it contains a brief future worth living at the end.

The trick, then, is to extend this analysis from futures to lives. But now a problem arises. In the case of futures, the implicit baseline is the life minus the future. But in the case of lives, we cannot say the baseline is the life minus the life. Removing the life from the life leaves nothing; the baseline vanishes. We’re back to the problem of the missing relatum. We could try some artificially constructed baseline such as, say, the empty set. But then we’d have to make sense of the *prima facie* bizarre idea that a life may be better or worse than the empty set. Or we could adopt an ontology of “absences” and say that the baseline is one of these. But absences are controversial; better to do without them if we can. My aim here is to propose solution that appeals neither to artificial constructions nor to absences.

In order to explain my proposal, it will be useful first to state my analysis of futures worth living more precisely. My aim is to do this in such a way as reveal a natural way in which to extend the analysis from futures to lives.

## 2.2. Formal Framework

First, I need to make a number of assumptions about the mereology and temporal structure of lives, as follows. Lives are temporally extended; they occur over periods of time. Each life is composed of temporal parts, which obey classical mereology. In particular, for any parts of a life, some part of the life is the unique “fusion” (or mereological sum) of those parts. Some parts are temporally *simple*, i.e. not composed of other temporal parts, and *instantaneous*, i.e. of no duration or temporal extension. Call these parts *slices*. Every slice of a life has a distinct temporal location, or *time*, and these times may be represented on a cardinal scale.<sup>1</sup> Every life has both a *first* (earliest) and a *last* (latest) slice, and these are distinct. The *duration* of a life is the difference in time between its first and last slice; so every life has some non-zero, finite duration.

If  $x$  is a life, then  $y$  is an *initial segment* of  $x$  if and only if (i)  $y$  is a part of  $x$ , and (ii) any slice of  $x$  which overlaps  $y$  is earlier than any slice of  $x$  which doesn’t overlap  $y$ . One life is a *truncation* of another if and only if the former is an intrinsic duplicate of an initial segment of the latter. For every life  $x$  and number  $t \in (0, 1]$ , there exists exactly one truncation of  $x$ , to be denoted by “ $x_t$ ”, the duration of which, expressed as a ratio of the duration of  $x$ , is  $t$ .<sup>2</sup> Thus, for example,  $x_{1/2}$  is the unique truncation of  $x$  that is precisely half as long as  $x$ , or, in other words, the unique whole-life intrinsic duplicate of the first half of  $x$ . We may therefore think of the truncation  $x_t$  as what would remain of  $x$  if it were “cut short” at time  $t$ , as  $x$  with its future, at  $t$ , removed. Since intrinsic duplication is a reflexive relation, every life is a truncation of itself; in particular,  $x = x_1$ .

Let  $X$  be the set of all possible lives, and let there be a “utility” function  $u : X \rightarrow \mathbb{R}$  such that  $u(x)$  represents the *value* or *goodness* of  $x$  on a *cardinal* scale.<sup>3</sup> This function may be determined by any theory of well-being you please, provided it is consistent with a cardinal scale, i.e. it allows us to make sense of ratios of differences of well-being. My proposal is intended to remain neutral, so far as possible, between such theories.<sup>4</sup>

We may then define for each life  $x$  a derived utility function  $u_x : (0, 1] \rightarrow \mathbb{R}$  such that  $u_x(t) = u(x_t)$ . This function represents a kind of temporal value:  $u_x(t)$  is the value of  $x$  at time  $t$ . It should be stressed, however, that this is *not* the value of the single, instantaneous slice of  $x$  that occurs at  $t$ , but rather the value of the whole initial segment of  $x$  that ends at  $t$ ; it is the value  $x$  *would* have were it to end at  $t$ .<sup>5</sup> Drawing this function on a graph shows

---

<sup>1</sup>The assumption of cardinality is stronger than required for my analysis; an ordinal scale would suffice. However, since a cardinal scale will be needed later, it will save complications to introduce it now.

<sup>2</sup>This implies that time is “continuous” rather than “discrete”. In this respect, the framework I adopt here differs from that adopted by John Broome in *Weighing Lives*. Broome assumes that time proceeds in “quantum steps” [Broome, 2004, p. 23]. This, he says, is merely a “modelling assumption”. For reasons that will become apparent, the perhaps simpler discrete-time model is not suitable for my purposes here.

<sup>3</sup>Technically, what I mean by this is that  $u$  is an arbitrarily selected member of a set of functions  $U$  such that  $U$ , and no non-empty proper subset of  $U$ , is closed under positive affine transformation. A *positive affine transformation* of  $u$  is a function  $u'$  such that, for some numbers  $\alpha$  and  $\beta$ , with  $\alpha > 0$ ,  $u'(x) = \alpha u(x) + \beta$  for every  $x$ , i.e.  $u'$  results from  $u$  by changing at most the *unit* and/or *zero*. Thus only those properties of  $u$  which are preserved under all such changes may be taken to have any significance. In particular, the *order* of the numbers  $u$  assigns to lives, and the *ratios of differences* between those numbers, may be significant; but, for example, whether the numbers are *positive* or *negative* cannot be.

<sup>4</sup>An alternative strategy would be to proceed on a case-by-case basis, giving a different definition of a life worth living for each theory of well-being. But merely doing this would not be to elucidate the *concept* of a life worth living. What we want to know is what all these different definitions must have in common.

<sup>5</sup>Strictly speaking, it is not the value of any part of  $x$ , but rather the value of a truncation of  $x$ , a distinct

the “shape” of the life, the way in which its value changes over time. An example is given in Figure 1. There is an initial period during which the value of the life is increasing, followed

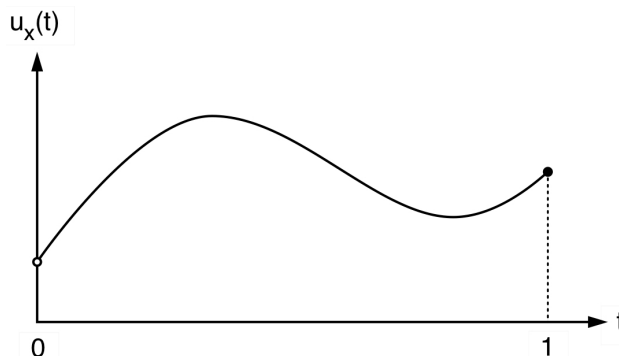


Figure 1: a possible life

by a period of decline, and then a final period of improvement at the end of the life. Note,  $u_x(0)$  is not defined, as indicated by the hollow dot at the left end of the curve. The reason is that there is no such truncation as “ $x_0$ ”. This would have to be a truncation whose duration was zero, but truncations are by definition whole lives and, as stated above, I’m assuming that all whole lives have some non-zero duration, that there are no instantaneous lives, in other words. Helping myself to instantaneous lives would make my task easier. As we’ll soon see, in their absence I need to rely on fairly strong assumptions about limits. Still, I find the notion of an instantaneous life too shaky to be relied upon. Of course I can’t deny that there are some instantaneous things; I’ve already assumed the existence of *slices*, instantaneous temporal parts of lives. What I’m reluctant to accept is that any of these things are *lives*. An analysis which depended on instantaneous lives would be, in my opinion, a version of the “artificial construction” strategy I set aside earlier. In this respect, an instantaneous life is like the empty set.

My proposed analysis of *futures* worth living may now be stated as follows.

**Futures.**  $x$  has a *future worth living* at  $t$  (with  $0 < t \leq 1$ ) iff  $u_x(t) < u_x(1)$ .

The future of  $x$  is worth living at  $t$  just in case the whole of  $x$  is better than that part of it which would remain if its future, at  $t$ , were removed. This is illustrated in Figure 2, using the same possible life as before. The value of the whole life is its value at the end, when  $t = 1$ , i.e.  $u_x(1)$ . Thus all and only those points on the curve which intersect the shaded region correspond to the times at which  $x$  has a future worth living. Moving from left to right, in the direction of time, we see the following. The curve begins within the shaded region, indicating an initial period during which all futures are worth living. It then climbs above the shaded region, indicating a period during which no futures are worth living. Finally, it dips down into the shaded region again, indicating a final period during all futures are worth living. In the middle period, it would be better for the life to end, rather than continuing. But in the surrounding initial and final periods, this is reversed: continuing is better than ending.

---

whole life which is an intrinsic duplicate of an initial segment of  $x$ . We can think of this value as being, in a derivative sense, the value of the initial segment itself. On the framework I’m developing here, however, all values are ultimately values of whole lives.

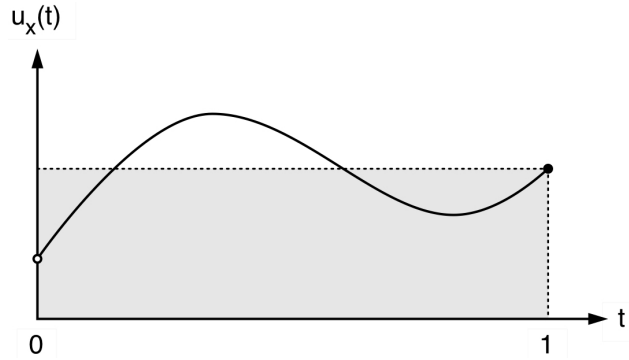


Figure 2: futures worth and not worth living

**2.3. From Futures to Lives**

This suggests a way of extending the analysis from futures to lives. Imagine beginning at some time  $t > 0$ , and then gradually moving backwards in time, making  $t$  gradually closer to 0. As  $t$  decreases, the future at  $t$  grows ever closer to being the whole life, and the corresponding baseline, the truncation  $x_t$ , shrinks ever closer to being nothing. We cannot actually have  $t = 0$ , so that the future is the whole life, because then the baseline would be lost. But, at least in favourable circumstances, we can do the next best thing: we can see what happens to the values of these ever-decreasing truncations as they *approach* nothingness, as  $t$  goes to 0. In the happy case, these values will “converge to a limit”, in the mathematical sense. We may then say that the life as a whole is worth living just in case this limit value is less than that of the whole life.

This yields the following analysis.

**Lives.**  $x$  is worth living iff  $\lim_{t \rightarrow 0} u_x(t) < u_x(1)$ .

Of course, this covers only what I’ve called the “happy case”, where the required limit exists. Nothing I’ve assumed so far guarantees that this will always be the case. In order for the analysis to be fully general, then, I need to make a further assumption.

**Limit Assumption.** For some  $\omega_x$ ,  $\lim_{t \rightarrow 0} u_x(t) = \omega_x$ .

The subscript in “ $\omega_x$ ” indicates that the limit is specific to the particular life  $x$ . Every life is assumed to have such a limit, but it needn’t be the same limit for all lives.

Is this a plausible assumption? Well, it’s surely not crazy. Let me sketch an argument for it. Consider two truncations of the a single life, one of which lasts for one millisecond and the other for two milliseconds. Any differences between these truncations must be extremely slight. One truncation is only a millisecond longer than the other, and very little can happen in a life in just one millisecond. On any plausible theory of well-being, any difference in *value* between the two must therefore also be extremely slight. It’s hard to imagine a theory of well-being according to which, say, a two-millisecond life might be vastly superior to a one-millisecond life. More generally, as the differences in duration between truncations become vanishingly small, the differences in value must become vanishingly small also. And that’s just to say, more or less, that their values must converge to a limit.

I don't deny that there could be sudden jumps in the value of a life. As others have suggested, people might gain well-being by successfully completing projects they have pursued. Think of life in which this happens, and consider two truncations, one ending slightly before the moment of success, and the other ending slightly after it. There might be a great difference in value between these two truncations, even though the difference in duration is very slight. However, I doubt that this could happen at the beginning of a life. You can't successfully complete a major project in, say, the first millisecond of your life.

So the Limit Assumption seems plausible at least for normal lives, ones we're accustomed to dealing with. If there are highly unusual lives for which the assumption doesn't hold, I might be prepared to say that it is indeterminate whether such lives are worth living (though I wouldn't want to commit to this in advance of considering some plausible examples).

### 3. An Alternative Proposal

I turn now to considering an alternative analysis suggested by John Broome [2004]. But I should, at the outset, enter a caveat: I shall fill out the details of the analysis in a way that fits the framework I've adopted here, and so I can't guarantee that Broome would endorse this particular way of developing his suggestions.

#### 3.1. Broome's Analysis

Broome distinguishes two senses of "a life worth living", a "temporal" sense and a "lifetime" sense, and defines the latter in terms of the former. About the temporal sense he says the following.

Suppose Elizabeth's way of life, as she is living now, is worth living. Living it is better for Elizabeth than not living it. Dying now would be worse for her than continuing to live. ... Elizabeth's life as a whole would be worse for her if she died than it would be if she continued to live. [Broome, 2004, p. 67]

This appears quite similar to my analysis of a *future* worth living. But there is, I think, a crucial difference. When Broome speaks of what Elizabeth's life would be like if she "continued to live", he means to refer, not necessarily to her *actual* future, but to the future she would have if her "way of life" remained the same, if her life kept going, as we might say, "in the same direction". Her *actual* future might not be like this; it might be that, actually, her life changes direction at some future time. So the future which, on Broome's account, is relevant in determining whether Elizabeth's *life* is worth living (in the temporal sense) need not be the same as that which, on my account, is relevant in determining whether her *future* is worth living. Whereas the latter is her actual future, the former may be some merely hypothetical future.

In the framework I've adopted here, the direction of a life at a time may be understood as the *rate of change* in the values of the truncations of the life at that time. For a life  $x$ , this is given by  $u'_x$ , the first derivative of  $u_x$  (assuming  $u_x$  to be differentiable). So the way  $x$  is going, or its *direction*, at  $t$  is given by  $u'_x(t)$ :  $x$  is going *well* or *badly* at  $t$  according as  $u'_x(t)$  is *positive* or *negative*. To illustrate, see Figure 3. This shows the rate of change in the value of

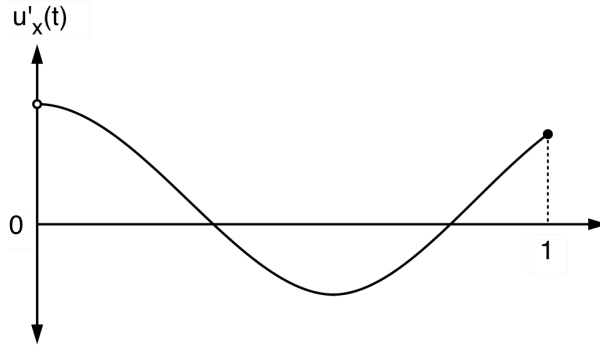


Figure 3: the direction of a life

the possible life discussed earlier (the one shown in Figures 1 and 2). It starts out going well, but the rate steadily declines and the life enters a period in which it is going badly. Then the rate increases again and the life enters a final period in which it's going well.<sup>6</sup>

Broome then moves from the temporal sense to the lifetime sense as follows.

Call a life 'neutral' at a time if it is just on the borderline between being worth living and not worth living at that time. Call a life 'constantly neutral' if it is neutral at every time. We might say that a life as a whole is worth living if and only if it is better than a constantly neutral life. [Broome, 2004, p. 68]

Developing this in the way suggested above, we may say that a life is *neutral* at a time if and only if its value is constant, getting neither better nor worse, at that time (i.e.  $x$  is neutral at  $t$  iff  $u'_x(t) = 0$ ); and it is *constantly neutral* if and only if its value is *always* constant. Thus we have the following alternative analysis of lives worth living.

**Lives (Broome).**  $x$  is worth living iff, for some  $x'$ ,

1.  $u'_{x'}(t) = 0$  for all  $t \in (0, 1]$ , i.e.  $x'$  is constantly neutral,
2.  $u(x') < u(x)$ .

This suggests a different interpretation of the implicit baseline. Here's an analogy. You're sharing a jug of beer with a group of friends. You go around the group pouring some beer from the jug into each person's glass. Being polite, you fill your own glass last, but there's only enough beer left to fill it half-way. "Oh well," you say, "at least that's better than nothing". In this case, we might say, the implicit baseline is an empty glass. A half-full glass is not as good as a full glass, but it's still better than an empty glass. Similarly, we might think of a constantly neutral life as an "empty" life, a life in which nothing happens.

To make the analogy more exact, suppose there's a function  $q : (0, 1] \rightarrow \mathbb{R}$  such that  $q(r)$  represents the quantity of beer (measured in litres, say) that is contained in that truncation of the glass whose volume, expressed as a ratio of the volume of the whole glass, is  $r$ . Thus,

<sup>6</sup>It can be seen that there are times at which the life is going well but its future is not worth living, and times at which it is not going well but its future is worth living. This should not be surprising. From the fact that a life is presently going well we cannot infer that it will continue to do so.

$q(1/2)$  is the quantity of beer in the bottom half of the glass. Now it's easy to see that the following equivalence holds:  $q(1) = 0$  if and only if  $q(r) = q(r')$  for all  $r$  and  $r'$  (or equivalently,  $q'(r) = 0$  for all  $r$ ). The whole glass is empty just in case every truncation contains the same quantity of beer (that quantity being zero). If the glass contains any (non-zero) amount of beer, then at least one truncation must contain more beer than at least one other truncation, and *vice versa*. Similarly, we might say that a life is empty, in the sense of containing zero well-being, just in case all its truncations contain the same quantity of well-being, that is, just in case all its truncations are equally good. This is equivalent to defining an empty life as a constantly neutral life.<sup>7</sup>

### 3.2. Comparing Broome's Analysis with Mine

I confess, I quite like Broome's analysis. I used to like it less because I found the temporal sense of "a life worth living" obscure. But now I see a way to understand this, I'm inclined to think the analysis is largely correct, or at any rate not obviously mistaken. Of course I also like my own analysis; it seems largely correct too. So I should hope that the two proposals turn out to be equivalent, at least given some plausible assumptions. As I aim to show, this is indeed the case.

To begin, we may see immediately a difference between the two analyses: only Broome's analysis presupposes the existence of constantly neutral lives. Or, more accurately, only Broome's analysis implies that if there are no constantly neutral lives, then no lives are worth living.<sup>8</sup> My proposal, on the other hand, is consistent with there being lives worth living without any neutral lives. But this seems a minor difference. We may eliminate it by making the following, fairly innocent, assumption.

**Neutrality Assumption.** At least one life is constantly neutral.

However, even given this assumption, the two proposals still might not be consistent. To see this, consider two lives,  $a$  and  $b$ , as illustrated in Figure 4. Clearly  $a$  is constantly

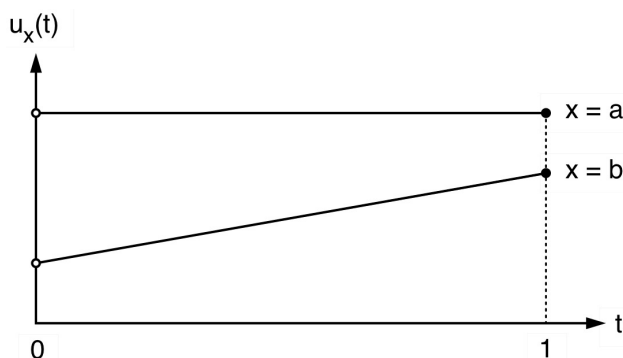


Figure 4: a problematic case

neutral. Suppose it's the only constantly neutral life. Then, since  $b$  is worse than  $a$ , Broome's proposal implies that  $b$  is not worth living. But the limit of the values of truncations of  $b$ , as

<sup>7</sup>We also have:  $q(1) > 0$  iff  $q(1) > \lim_{r \rightarrow 0} q(r)$ . So this analogy fits nicely with my analysis too.

<sup>8</sup>Of course by "lives" I here mean *possible* lives. Broome's proposal doesn't require that any *actual* lives are constantly neutral.

those truncations tend to nothingness, is clearly less than the value of the whole of  $b$ . Thus, on my proposal,  $b$  is worth living. So we have a contradiction. If pairs of lives like  $a$  and  $b$  are possible, the two proposals are inconsistent.

We might then ask which of the two proposals is more plausible *in this case* where they disagree. The answer, I think, is that both are implausible and roughly equally so. Consider two claims:

**Claim 1.** If a life is *always* going well, then the life as a whole is worth living. If a life is *never* going well, then it is not worth living.

**Claim 2.** A life that is worth living is better than a life that is not worth living.

These two claims are both very plausible.<sup>9</sup> But together they are inconsistent with the possibility of pairs of lives like  $a$  and  $b$ . Although  $b$  is always going well and  $a$  is never going well,  $a$  is better than  $b$ ; and this is clearly impossible given the two claims. Therefore, if we allow the possibility of such pairs of lives, we must reject one of the two claims. Broome's analysis rejects Claim 1; it says  $b$  is not worth living even though it's always going well. My analysis rejects Claim 2; it says only  $b$  is worth living even though it's worse than  $a$ . Neither rejection is a happy one.

A better alternative might be to reject the possibility of such pairs of lives. One way to do this is to assume the following.

**Convergence Assumption.** For some  $\omega$ ,  $\lim_{t \rightarrow 0} u_x(t) = \omega$ .

This goes further than the Limit Assumption by assuming, not only does every life have a limit, but the limit is the same for all lives (hence " $\omega$ " has no subscript). Roughly speaking, the assumption is that every life begins in the same place. This rules out lives  $a$  and  $b$  because, as the graph shows, they begin in different places;  $a$  starts higher up than  $b$ .

The argument I gave earlier in support of the Limit Assumption also supports the Convergence Assumption. Very little of significance can happen in a life in a very short space of time. Thus, as the differences in duration between truncations of lives become vanishingly small, the differences in value between them must become vanishingly small also; and this will be so even if they are truncations of *different* lives.

Given the Convergence Assumption and the Neutrality Assumption, Broome's proposal and mine are *equivalent*, and they imply both Claim 1 and Claim 2. Thus, setting aside the Neutrality Assumption, the fortunes of the two proposals seem very closely wedded: if the Convergence Assumption is false, then neither proposal should be accepted; but if it is true, then we cannot consistently accept one without the other. In the unlikely event that the Convergence Assumption is true and the Neutrality Assumption false, only my proposal remains acceptable (unless we're prepared to say that no lives are worth living). This might give my proposal an edge, but only a slight one.

---

<sup>9</sup>One way in which to reject Claim 1 would be to reject the very notion of a "life worth living", to claim that no life is worth living, not because all lives are terribly bad, but because talk of lives worth living is meaningless or incoherent. This would come at a cost. As I say, this notion is ubiquitous in population ethics, and it's not clear that it could easily be dispensed with. Though it should be noted that Broome himself is happy to do without it. Broome of course can't deny the coherence of the notion, since he proposes an analysis of it. Nonetheless, he claims to find it unhelpful and eschews any use of it in developing his account of population ethics. See Broome [2004, p. 68].

## 4. Some Implications

I shall finish by noting some implications of the above discussion for the task of comparing lives of different duration. It will be helpful to assume a particular, substantive theory of well-being, namely, *hedonism*, the theory that well-being is pleasure. But my conclusions may apply more generally to other theories of well-being.

### 4.1. Hedonism: Simple and General

Perhaps the simplest form of hedonism says that the value of a life is given by the *total* quantity of pleasure it contains. (For simplicity, I'll set aside pain.) This may be defined more precisely as follows. Let there be a function  $p : X \rightarrow \mathbb{R}$  such that  $p(x)$  represents the total quantity of pleasure contained in  $x$  on a ratio scale, where zero represents the complete absence of pleasure. Then we have the following theory.

**Simple Hedonism.**  $u(x) = p(x)$ .

In this case, it is plausible that the Convergence Assumption holds, and, in particular, that  $\lim_{t \rightarrow 0} u_x(t) = 0$  for every  $x$  (i.e. roughly, every life begins containing no pleasure). If so, then combining Simple Hedonism with my proposal yields the expected result that  $x$  is worth living if and only if  $p(x) > 0$ ; a life is worth living just in case it contains *some* amount of pleasure. This seems to be the view of lives worth living that one should accept if one accepts Simple Hedonism.

However, this simple form of hedonism faces an obvious difficulty, a variant of Derek Parfit's "Repugnant Conclusion". A fixed total of pleasure may be contained in lives of varying duration. It may be concentrated in a short life, or spread thinly in a long life. An extremely long life containing only very mild pleasures may contain the same total as a shorter life containing more intense pleasures. Yet it is implausible to say that any two such lives must be equally good, as this simple hedonistic view implies.

A natural way to solve this problem is to adopt a form of hedonism that takes into account, not only total pleasure, but also *average* pleasure (per unit of duration). This may be done as follows. Let there is a function  $d : X \rightarrow \mathbb{R}$  such that  $d(x)$  represents the duration of  $x$  on a ratio scale, in the natural way. So we have  $d(x_t) = td(x)$ . Now we may define the following family of hedonistic theories of well-being.

**General Hedonism.** For some  $r \in [0, 1]$ ,

$$u(x) = \frac{p(x)}{d(x)^r}$$

Each member of the family assigns a different value to the parameter  $r$ , which determines the relative weights given to the total and the average pleasure contained in a life. At the two extremes, only one of total or average counts. When  $r = 0$ , the value of a life is just the total pleasure it contains. (So Simple Hedonism is a special instance of General Hedonism.) When  $r = 1$ , the value of a life is just the average pleasure it contains. In intermediate cases, both total and average are given some weight, more or less depending on how close  $r$  is to either 0 or 1.

## 4.2. Problems for General Hedonism

General Hedonism avoids the version of the Repugnant Conclusion discussed above, but of course only if  $r > 0$ . However, it is then inconsistent with either Claim 1 or Claim 2, given the following rather weak condition.

**Non-arbitrary Domain.** For every  $r > 0$ , there are lives  $a$  and  $b$ , and positive numbers  $\alpha$  and  $\beta$  such that:

1.  $d(a) = d(b)$ ,
2.  $p(a_t) = \alpha t^r$ ,
3.  $p(b_t) = \beta t^{r+1}$ ,
4.  $\alpha > \beta$ .

To see this, let  $a$  and  $b$  be as defined in Non-Arbitrary Domain. Then General Hedonism implies that

$$u_a(t) = \frac{p(a_t)}{d(a_t)^r} = \frac{\alpha t^r}{(td(a))^r} = \frac{\alpha}{d(a)^r},$$

and that

$$u_b(t) = \frac{p(b_t)}{d(b_t)^r} = \frac{\beta t^{r+1}}{(td(b))^r} = \frac{\beta t}{d(b)^r}.$$

Notice,  $u_a$  is a constant function, but  $u_b$  is strictly increasing (i.e. if  $t < t'$ ,  $u_b(t) < u_b(t')$ ). So  $a$  is never going well and  $b$  is always going well. Claim 1 then implies that  $a$  is not worth living and  $b$  is worth living. But, since  $d(a) = d(b)$  and  $\alpha > \beta$ , we have

$$u_a(1) = \frac{\alpha}{d(a)^r} > \frac{\beta}{d(b)^r} = u_b(1).$$

So  $a$  is better than  $b$ . Claim 2 then implies that  $b$  is worth living only if  $a$  is worth living. Thus we have a contradiction.

If two lives such as these are possible, then General Hedonism is in trouble. But, so long as  $r > 0$ , I see no reason to think all such pairs of lives are impossible.<sup>10</sup> This seems, therefore, to be quite a serious objection to General Hedonism: either General Hedonism implies a version of the Repugnant Conclusion, or it is inconsistent with the conjunction of Claim 1 and Claim 2. Of course this is far from saying that *all* forms of hedonism are to be rejected. Nonetheless, it is interesting, I think, that consideration of the concept of a life worth living may lead us to reject one family of hedonistic theories.

## References

JOHN BROOME [2004], *Weighing Lives*, Oxford University Press.

---

<sup>10</sup>If  $r = 0$ , then the definition of  $a$  violates the earlier assumption that every life begins containing no pleasure.