

Stockholm University  
Department of Philosophy

## Wellbeing Gaps and Repugnant Conclusions

Luca Stroppa

E-mail Address: [quelo42@gmail.com](mailto:quelo42@gmail.com)

T-Number: 950921T273

Master's thesis in Practical Philosophy

30 credits

Spring term 2019

Supervisor: Krister Bykvist

## TABLE OF CONTENTS

SECTION 1. INTRODUCTION.....	3
1.1 GOLDEN AND IRON AGE .....	3
1.2 DEFINITIONS AND BACKGROUND ASSUMPTIONS.....	3
1.3 OVERVIEW OF THE THESIS .....	5
SECTION 2. RC: FOR AND AGAINST .....	6
2.1 REVISING AN INTUITION .....	6
2.2 THE APPEALING RC'S PREMISES.....	6
2.3 EQUIVALENCE AND MORE-IS-BETTER: HOW HUEMER MISSES QC .....	8
2.4 HUEMER'S BIAS ANALYSIS.....	10
2.5 TÄNNSJÖ AND RYBERG: LIVING IN PRIVILEGED SOCIETIES .....	14
2.5.1 Tännsjö's argument.....	14
2.5.2 Ryberg's preference satisfaction view .....	17
2.5.3 Ryberg's and Tännsjö's impressionistic evaluations .....	18
2.5.4 Sufficiently challenging QC .....	21
2.6 REVERSED RC AND REQUIREMENTS FOR A THEORY.....	26
SECTION 3. A GAP FOR THREE WELLBEING THEORIES.....	29
3.1 SATISFYING THE REQUIREMENTS .....	29
3.1.1 Two common features.....	29
3.1.2 Wellbeing gaps.....	33
3.1.3 The structure of my argument .....	36
3.2 MENTAL STATE THEORIES.....	38
Step 1: Mental state theories.....	38
Step 2: Wellbeing hypersensitivity denial .....	39
Step 3: Life Z.....	40
Step 4: Life A.....	40
Step 5: Comparison .....	40
Step 6: RRC.....	41
3.3 PREFERENCE SATISFACTION THEORIES .....	41
Step 1: Preference satisfaction theories .....	41
Step 2: Wellbeing hypersensitivity denial .....	42
Step 3: Life Z.....	42
Step 4: Life A.....	44
Step 5: Comparison .....	44
Step 6: RRC.....	44
3.4 OBJECTIVE LIST THEORY .....	44
Step 1: Objective list theories .....	44
Step 3, here 2: Life Z.....	45
Step 2, here 3: Hypersensitivity denial.....	46
Step 4: Life A.....	47
Step 5: Comparison .....	47
Step 6: RRC.....	47
SECTION 4. CONCLUSION .....	48
BIBLIOGRAPHY.....	49

## Section 1. Introduction

### 1.1 Golden and Iron Age

In his poem *The Works and Days* Hesiod describes the five ages of Greek Mythology. Chronologically, the Golden Age is the first. Back then humans lived without sorrow or grief and earth provided them everything they could need. It is unclear how many humans lived the golden age, but the number of people was presumably much smaller than it is today.<sup>1</sup>

Hesiod instead, and us with him, is collocated in the last age, the Iron age. Nowadays mankind knows the pain from toil and sorrow, and fatigue is necessary for surviving. On the other hand, humans are much more in number.<sup>2</sup>

Plausibly in our age, despite our labours and sorrows, our lives are on average worth living. Since humanity is way larger now than in the Golden Age, the total human wellbeing might be higher now than back then. Is our population better or worse than the population living in the Golden Age with respect of wellbeing?

I argue in this work that the current population might actually be better.

### 1.2 Definitions and background assumptions

Population Ethics is the field of moral reasoning that seeks “an adequate theory of population value where the number of people, their welfare, and their identities may vary” (Arrhenius & Campbell, 2017).

A population is defined as a set of individual lives. Such lives might coexist or exist at different times. It is assumed that someone’s life can be better, worse, or as good as someone else’s life with respect of wellbeing; unless where otherwise stated, I assume full comparability in this work. I assume that a life can have positive, negative, or neutral wellbeing (or undefined wellbeing). The wellbeing of a life is positive iff the life is worth living and the wellbeing of a life is negative iff the life is worth not living. In this work, I use the expressions “wellbeing” and “welfare” interchangeably. For the sake of simplicity, I assume that each life has the same duration. Without this assumption the discussion would be needlessly complex.<sup>3</sup>

This thesis is mainly focused on a discussion of

---

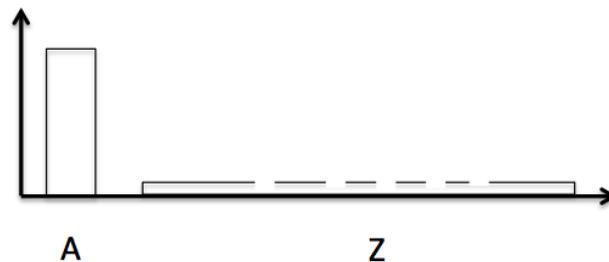
<sup>1</sup> See (Hesiod, 700 BC, verses 109-120)

<sup>2</sup> See (Hesiod, 700 BC, verses 174-201)

<sup>3</sup> Consider for example the following. Let A, B and C be lives with positive wellbeing. A’s duration is as long as the sum of B and C’s durations, and A contains as much positive wellbeing as B and C combined. Is it better that A exist while B and C do not exist, or that B and C exist and A never exist? Should we value, for example, longevity over variety of lives? Problems concerning lives with different durations deserve more attention than they have received so far given in the literature but I will not examine them here.

*The Repugnant Conclusion* (RC): For each population of very many people— say, ten billion—all of whom have a very high quality of life, there must be some much larger number of people whose existence, would be better with respect to wellbeing, even though these people would have lives that are barely worth living. (Parfit, 2004, p. 10)

An illustration might make RC clearer. Call A the population with very many lives who have very high welfare and Z the population with many more lives, but these lives are barely worth living. The width of the rectangles is the number of lives, their height is the level of welfare. Population Z has dashes to signify that Z can be arbitrarily large. The horizontal arrow is the neutrality level: positive levels of wellbeing are above it.



I ask to the reader to remember these features of population A and population Z, since I will frequently refer to them in this work.

A theory for Population Ethics avoids RC if it does not imply that some large population with lives at high welfare is worse than any much larger population with way lower welfare. In the literature, the intuitive principle of RC's avoidance has been identified with:<sup>4</sup>

*The Quality Condition* (QC): There is at least one population with very high positive [average] welfare which is better with respect to wellbeing than any population with very low positive [average] welfare. (Arrhenius, 2000, p. 41)

In other words, any amount of lives at a very low positive welfare Y is worse than some numbers of lives at a very high positive number X. This relationship between X and Y is often described in the literature with the sentence: "X is weakly value-superior to Y".<sup>5</sup> Weak

---

<sup>4</sup> The logical relation between QC and RC is formally demonstrated in (Arrhenius, 2000, pp 205-206). I consider this relation sufficiently clear and do not report it in this work.

<sup>5</sup> See for example (Arrhenius, 2005, pg. 98).

value-superiority is distinguished from strong value-superiority: X is strongly value-superior to Y iff any amount of X is better than any amount of Y.

QC is an essential principle in the several impossibility theorems, presented by Arrhenius,<sup>6</sup> that show that a theory for Population Ethics cannot satisfy some sets of very appealing conditions.

RC is often used as an objection against the following principle for deciding which population is better:<sup>7</sup>

*The Total Principle:* A population X is better with respect to wellbeing than a population Y if and only if the total amount of wellbeing is greater in X than in Y (Carlson, 1998, p. 291).<sup>8</sup>

The clause “with respect to wellbeing” means that the *Total Principle* as considered in this work does not make any claim on whether the value of a population is influenced by fairness, desert, equality, economic features or similar features *per se*. Again, an inquiry on this matter would be important for Population Ethics, but it will not be a concern of this work.

If RC is true, then QC is false. If QC is false, then the *Total Principle* is supported. In this thesis I try to find the features of a wellbeing theory that would make less problematic to accept RC. If it is plausible for wellbeing theories to have these features, which there is reason to think, then our confidence in QC should be reduced and our confidence in the *Total Principle* strengthened.

### 1.3 Overview of the thesis

In Section 2, I examine Michael Huemer’s, Torbjörn Tännsjö’s and Jesper Ryberg’s arguments in defence of RC, that is to say, against QC. In this section I present their arguments and their replies to some objections. I also provide them further objections, showing why they fail to defend RC. Furthermore I point out a limit of a strong objection by Arrhenius to Tännsjö and Ryberg. I conclude the section with a list of features that a wellbeing theory should have in order to support the *Total Principle*.

In Section 3, I try to find this wellbeing theory. In the first part of the section I propose a way to satisfy the features listed at the end of section 2. The crucial element of my proposal

---

<sup>6</sup> Six of them can be found in (Arrhenius, 2000, p. 151-198), others can be found in (Arrhenius, 2003) and (Arrhenius, 2017).

<sup>7</sup> How the *Total Principle* implies RC is shown in (Arrhenius, 2000, pg. 49-50)

<sup>8</sup> See also (Parfit, 1984, p. 387)

is introducing a gap in the ordering of wellbeing levels between the level where life is barely worth living and the level where life is barely worth not living. I provide arguments for the plausibility of this gap throughout the whole section. In the second part of the section I examine at least one theory of wellbeing for each kind in Parfit's classical taxonomy: I analyse some mental state theories, a preference satisfaction theory and an objective list theory. I enquire whether these theories can plausibly allow the required gap and whether RC conflicts less with our intuitions if they do. I find a positive answer for both.

Section 4 is a conclusion of the text with a summary of the work.

## Section 2. RC: for and against

### 2.1 Revising an intuition

Huemer uses, in (Huemer, 2008), all the main arguments defending RC written to date. Huemer's aim is to show that the intuition of RC's repugnance should not be trusted. In other words Huemer is challenging QC, even if he never explicitly recognizes its existence.

I summarize in 2.2 Huemer's demonstration of how RC follows from three very appealing premises. In 2.3 I focus on further arguments for accepting RC, called The Equivalence Argument and the More-Is-Better Argument. I also show how, in such arguments, Huemer fails to challenge satisfactorily QC. Then I present Huemer's bias analysis in 2.4. This analysis is, in my opinion, the strongest part of Huemer's defence of RC. In 2.5 I explain how Ryberg and Tännsjö strengthen the bias analysis according to which we tend to underrate low-quality lives and I criticize their impressionistic account of lives barely worth living. In 2.6 I show that, even if one accepts RC, there are further problems for Huemer's, Tännsjö's and Ryberg's theories. I conclude the section by showing what would be required in order to defend the *Total Principle* against RC and the problems highlighted in 2.6.

### 2.2 The Appealing RC's Premises

Huemer's first argument against RC's Repugnance is called the Benign Addition Argument.<sup>9</sup> The point of the argument is to show how RC is implied by the following very appealing premises:

---

<sup>9</sup> The Benign Addition Argument is actually one of the many variations of the Mere Addition Paradox. The first three versions of Mere Addition Paradox can be found in (Parfit, 1984, pg. 419-442) and two further versions by Parfit are in (Parfit, 2004). Huemer's proof here summarized is clearly inspired by (Parfit, 2004, pp. 14-17). Interesting variations can be found also in (Rachels, 2004).

*The Benign Addition Principle*: If worlds  $x$  and  $y$  are so related that  $x$  would be the result of increasing the well-being of everyone in  $y$  by some amount and adding some new people with worthwhile lives, then  $x$  is better than  $y$  with respect to wellbeing.<sup>10</sup>

*Non-anti-egalitarianism*: If  $x$  and  $y$  have the same population [size], but  $x$  has a higher average wellbeing, a higher total wellbeing, and a more equal distribution of wellbeing than  $y$ , then  $x$  is better than  $y$  with respect to wellbeing.<sup>11</sup>

*Transitivity*: If  $x$  is better than  $y$  with respect to wellbeing and  $y$  is better than  $z$  with respect to wellbeing, then  $x$  is better than  $z$  with respect to wellbeing.<sup>12</sup> (Huemer, 2008, pg. 903).

The proof goes as follows. Huemer describes three worlds:

*World A*: One million very happy people (welfare level 100).

*World A+*: The same one million people, slightly happier (welfare level 101), plus 99 million new people with lives barely worth living (welfare level 1).

*World Z*: The same 100 million people as in A+, but all with lives slightly better than the worse-off group in A+ (welfare level 3). (Huemer, 2008, p. 903)

Since A's million people would be better off in A+ and there are further lives worth living, A+ has to be preferred to A from *The Benign Addition Principle*. Since A+ and Z have the same population, but Z is better than A+ as for average wellbeing (Z has 3, A+ has slightly more than 1), total wellbeing (Z has 300 million, against the 100 million and hundred) and equality of welfare's distribution, then Z has to be preferred to A+ due to *Non-Anti-Egalitarianism*. Since Z has to be preferred to A+ and A+ has to be preferred to A, *Transitivity* implies that Z has to be preferred to A, which is RC.

Huemer tries to demonstrate that these premises are uncontroversial in (Huemer, 2008, pp. 904-907).<sup>13</sup> However, RC's opposers would still insist that a theory for Population Ethics implying RC strikes them as unacceptable, and many prefer to abandon one of the three

---

<sup>10</sup> See note 11. The Benign Addition Principle can be seen as the conjunction of Arrhenius' Quantity Condition (Arrhenius, 2000, pg. 155-156) and Egalitarian Dominance Condition (Arrhenius, 2000, pg. 156-157)

<sup>11</sup> In the original formulation the term "wellbeing" is replaced with "utility". It is unclear to me why Huemer mentions equality once he puts the clause "with respect to utility".

<sup>12</sup> See note 11.

<sup>13</sup> His demonstrations received some challenge. For example, see (Goodrich, 2016).

premises rather than QC.<sup>14</sup> This is so because QC, indubitable to some thinkers, is not very much challenged by the Benign Addition Argument.

In order to strengthen his argument Huemer lists, in (Huemer, 2008, pp. 911-915), a great number of attempts to formulate a theory of Population Ethics that avoids RC, and shows how each attempt has implications even more problematic than RC. This is not surprising, given that, as I pointed out in the introduction, Arrhenius proved seven impossibility theorems in Population Ethics.

Unfortunately the consideration concerning failed attempts, once again, does not challenge QC at all. Furthermore, the possibility of a theory of Population Ethics that avoids RC and whose implications are less problematic than RC is still open.

The fact that the three appealing premises imply RC and that, so far, any attempt to avoid RC has even more problematic implications is not sufficient to reject QC. But, as we will see in the next section, Huemer adds two further arguments for the acceptance of RC.

### 2.3 Equivalence and More-is-Better: how Huemer misses QC

The first argument is called the Equivalence Argument. It gets its name from, and is based on, the relation “ $x$  is at least equivalent to  $y$ ”, which is reflexive and transitive and means that an increase in  $x$  is at least as good with respect to wellbeing as a proportionate increase in  $y$ . Its first premise is:

1. *Any duration of a benefit is at least equivalent to some increase in intensity of benefit.*

By ‘increase in intensity’ of a benefit Huemer means how much someone’s level of wellbeing is increased during the time she enjoys a benefit. The premise can be restated as follows: experiencing any increase  $x$  in duration of a benefit is at least as good, with respect to wellbeing, as experiencing some increase  $y$  in intensity of a benefit (not necessarily the same benefit). For example, the increase of wellbeing given by the benefits of watching an Enya music video and, later, eating a sandwich seems at least as good as the increase of wellbeing given by the two benefits simultaneously.

The second premise of the Equivalence Argument is:

2. *Any number of people experiencing a benefit is at least equivalent to some duration of a*

---

<sup>14</sup> A clear example of this will come from Parfit, that defended Perfectionism, stating that it “seems to [him], at times, crazy. But [...] without Perfectionism how can we avoid the Repugnant Conclusion?” (Parfit D., 2004, pg. 20). Perfectionism violates *Non-anti-egalitarianism*.

*benefit.*<sup>15</sup>

In other words, a situation where a person P is experiencing benefits  $x$  and, later, P will experience  $y$  is at least as good as a situation where a person P is experiencing benefit  $x$  and a person  $P_1$  sufficiently similar to P is experiencing benefit  $y$ . That is to say, enabling P to eat a sandwich today and watching Enya tomorrow is at least as good as allowing P to eat the sandwich today and allowing  $P_1$  to watch Enya tomorrow.

Transitivity and the two premises imply that any number of people experiencing a benefit is at least as good with respect to wellbeing as some intensity of benefit. In Population Ethics terminology, any increase in the number of people with positive welfare is at least as good as some increase in average welfare. Thus a decrease in average welfare can be compensated with respect to wellbeing by an increase in number of people with positive welfare. This implies RC.

The second argument is called the More-Is-Better Argument. The argument is rather straightforward: lives with positive welfare are good, and more of a good thing is better, thus more lives with positive welfare are better. Since, according to Huemer, the marginal value of additional lives with positive welfare does not diminish, adding any number of lives with positive welfare is better. Thus, a sufficiently large number of lives with positive welfare, even if such welfare is very low, is better than a number of lives with high positive welfare: that is, RC.

I do not believe that the two arguments are sufficient to abandon QC. The point of QC is that there seems to be a level of welfare X that is weakly value-superior than a lower welfare Y, that is to say, any amount of lives at a very low positive welfare Y is worse than some numbers of lives at a very high positive number X. The first premise of the Equivalence Argument is that any increase or decrease in duration of experienced welfare has some equivalent in an increase or decrease in intensity of welfare, that is to say, any high welfare experienced for a duration  $x$  is equivalent to some lower welfare experienced for a duration  $y$  longer than  $x$ . This premise is what Huemer needs to demonstrate with the Equivalence Argument: this premise alone implies a denial of QC. Since QC's negation is implied by this premise, the argument fails to challenge QC.

Similarly, the More-Is-Better Argument shows that adding lives with positive welfare is always good, but in order to challenge QC it needs to show, in a plausible way, that a

---

<sup>15</sup> I assume that Huemer believes this condition is uncontroversial as long as prolonging the duration of benefit does not entail prolonging a life.

decrease in average welfare can be compensated by a sufficient increase in people with positive welfare even in the extreme case represented by RC. Furthermore, there are theories in which RC is avoided even if it is always good to have more people with worthwhile lives. For example, (Carlson, 2010) introduces a theory according to which (1) a population is better with respect to wellbeing if people with any kind of positive welfare are added, (2) there is no decreasing marginal value in adding people with positive welfare, but (3) no amount of people with very low positive welfare is equal in value to any number of people enjoying very high welfare, as prescribed by QC. Since QC is compatible with the conclusion of the More-Is-Better Argument, the argument is insufficient to decrease our confidence in QC.

However, Huemer points out that QC is not supported by any arguments, but only by a bare appeal to intuition. For this reason I think that his most successful set of arguments against QC is the list of biases that might distort our judgment in favour of QC. In fact, if there is no argument supporting QC and many explanations of why our intuition might be distorted in favour of QC, then we have a reason to distrust it. However, as I will show, it will be not a sufficient reason to abandon QC.

#### 2.4 Huemer's bias analysis

The first bias Huemer considers is *the egoistic bias* (Huemer, 2008, p. 907). RC suggests preferring a world with a very large population in which lives are barely worth living rather than a world with a smaller population in which lives enjoy high welfare. In evaluating RC's appeal, it is possible that people ask themselves *not* which population is best, *but* in which one they would like to live. Of course, people prefer enjoying a higher welfare, thus the world suggested by RC seems unappealing. But the population in which we would prefer to live is not necessarily the best population: in considering RC's correctness, we are biased against RC by a question that is easier to answer but has nothing to do with the original question. In fact, answering to the question "would you prefer living in a world in which you have higher welfare or a world in which you have low, but positive, welfare?" is easier than answering the question "is a world with higher average wellbeing better with respect to wellbeing than a world with higher total wellbeing?".

This argument by Huemer can be further strengthened by Kahneman's research on biases: in (Kahneman, 2011, pp. 96-104) he demonstrates empirically how people tend to answer a complex question by substituting it with an easier question. I am, however, not convinced of the relevance of *the egoistic bias* in Population Ethics: I do not believe that *all*

RC opposers are biased in this sense. This means that getting rid of this bias is insufficient to make RC acceptable.

Huemer then considers *The large number bias* (Huemer, 2008, p. 907-908). People struggle at grasping very large numbers. Consider the following statements:

- the distance from Earth to Betelgeuse is approximately 6 trillions of kilometres
- the distance from Earth to Betelgeuse is approximately 6 thousand trillions of kilometres

If we try to imagine in our head the two situations, we visualize distances that are more or less alike, even if the former is a thousand times bigger than the latter. Similarly, imagine a population of 10 billion people with very high welfare, which is population A in RC. The number of people in Z will be much higher, say, a billion of billions. The large number bias might impede to correctly appreciate how many more happy lives are there in Z, thus making RC less appealing than it should.<sup>16</sup>

According to Huemer, people are also biased while *Compounding small numbers* (Huemer, 2008, p. 909-910). Huemer lists some examples in which people make mistakes when they compound very small quantities. I report what is, in my opinion, the most striking one.<sup>17</sup>

Suppose the existence of a very large piece of paper that is just one hundredth of millimetre thin. If the paper is folded in half, it will be two hundredth of millimetres thin. If the piece of paper's thickness doubles each time it's folded, how thick will it be after being folded fifty times? The intuition of many people will lead to believe that the piece of paper will be a couple hundred meters thick, but it will be actually more than ten million kilometres thick.<sup>18</sup>

---

<sup>16</sup> Huemer does not use distance but a longer, perfectly equivalent example with time. Notice however that in Huemer's example, and in my analogous example, we fail to appreciate the difference in quantity of the same parameter, namely, distance or time. This is not perfectly analogous to RC where, if we have a *large number bias*, would have problems in estimating the amount of total good given that we cannot estimate correctly the number of people. This observation holds also for Mogensen's observation on a Korea war, discussed in this section at page 13/14.

<sup>17</sup> The curious reader will find such examples at page (Huemer, 2008, pg. 909-910).

<sup>18</sup>  $1 \times 10^{-5} \text{ m} \times (2^{50}) = 112 \times 10^8 \text{ m} = 11,2 \times 10^6 \text{ km}$ . Huemer's original example is about inches. Krister Bykvist suggested to me that, in the case of this example, we might have wrong intuitions mainly because the piece of paper here considered is an ideal object rather than a real one. A more realistic, but less striking Huemer's example is that people tend to express greater willingness to wear a seatbelt after reading report about their *lifetime* risk of being injured in a traffic accident rather than after reading reports about their risk *per trip* of being injured in a traffic accident. I find this example less convincing because the math behind summing risks is less intuitive than than the math for summing quantities. It is not a simple addition: if I flip a coin I have 50% probability that it will land showing head, but if I throw it twice the probability of head is just 75%, not 100%.

If our intuition fails to appreciate the sum of small values, then it probably also fails to appreciate why *Z* is more appealing than *A*. It is, however, unclear to me what the difference is between *the large number bias* and the *compounding small numbers bias*, given that large numbers are sums of small numbers.

Furthermore, Huemer notices that we might suffer from the *The actualist bias* (Huemer, 2008, 915-917).<sup>19</sup> An external observer might advise a couple with children against having another child, because it would lower the wellbeing of already existing children. Once a child is born, instead, the external observer does not regret the child's existence. An observer with a retrospective attitude allows the lowering of the average wellbeing in order to have an additional person. In contrast, an observer with a prospective attitude seems to be against it. This is presumably because an observer with a retrospective attitude fails to realize that people do not actually regret lowering the wellbeing of present people in order to add people with worthwhile lives. An observer with a prospective attitude is biased against non-existing people. This leads to an unjustified reluctance in lowering wellbeing in order to create new happy people. Thus, our intuition of RC's repugnance might be suggested by the *actualist bias*.

Huemer's claim on the existence of an *actualist bias* can be strengthened by referring to studies of experimental psychology such as (Kahneman, 2011, pp. 84-88). According to this study people tend to "fail to allow for the possibility that evidence that should be critical to our judgment is missing—what we see is all there is" (Kahneman, 2011, p. 88): for example, people tend to answer "yes" to the question "do you think that Rob, who is intelligent and strong, would be a good leader?", failing to realize that Rob might also be cruel and corrupt. In the same way a person with prospective attitude might be failing to correctly consider the value of the existence of an additional person. A person with retrospective attitude might instead make more reliable considerations on the additional person's existence.

However, even taking into account Kahneman's research, I am not certain that we are affected by *the actualist bias*. Huemer needs an argument for convincing me that we should trust the retrospective attitude rather than the prospective attitude. It is true that an observer with the prospective attitude might be failing to correctly take into account a future child's value, but this observer might be failing to correctly take into account the value of the loss in welfare experienced by the other two children. In other words, an observer with retrospective

---

Thus the different attitude of the seatbelt example might be due to different difficulties in understanding what a statistic means: it is clearer to understand the impact of a risk per lifetime rather than per episode.

<sup>19</sup> Despite the name, in (Huemer, 2008, 915-917) the *actualist bias* is presented as a stand-alone argument against QC, not in the list of biases.

attitude might be biased as well, since this observer does not consider the counterfactual world where the new child does not exist.

The most important bias for the purposes of this thesis is the one concerning *Underrating low-quality lives* (Huemer, 2008, p. 910). Huemer claims that it is unclear what a life barely worth living is, or how it would look like. In fact, how much a life is worth living depends probably on a complex balance of goods and bads of different kind. Since this is so, people might mistake a life barely worth living with a life worth not living. For example, Hudson believes that lives barely worth living are lives of people “not doing anything interesting but merely struggling for a bare subsistence”.<sup>20</sup> Parfit thinks that a life barely worth living is, though worth living, “gravely deficient in all of the features that can make a life worth living” (Parfit, 1984, p. 433). But formulation such as Hudson’s and Parfit’s give the wrong suggestion: they point out a negative feature of a life barely worth living, without specifying why such a life still remains worth living. Without this specification, their formulation could be fit for a life worth not living as well as for a life barely worth living. Since it is unclear what a life barely worth living should look like, RC is mistakenly represented as repugnant.

I think that Huemer plausibly shows that, due to some biases, world Z seems worse than it actually is.<sup>21</sup> This is, once again, not enough to abandon QC and thus accept RC: what Huemer needs to show is that world Z is better than world A, not just that the wellbeing level of world Z is not that unappealing. Despite any bias we could have, our intuition that A is better than Z might still be reliable. Taking *the large number bias* as an example, as pointed out in (Mogensen, forthcoming), from our simple incapability to imagine very large numbers it is hard to infer our ineptitude to morally evaluate large populations.

Borrowing Mogensen’s example, suppose that a Korea war is inevitable and there can be 5 million or 5 billion deaths. When I try to imagine 5 million casualties and 5 billion casualties, I might fail to appreciate precisely how much more horrendous the latter is than the former. Nevertheless, I would *still* know that 5 billion casualties are morally worse than 5 million casualties: even if my gut reaction does not fully appreciate the difference in badness, I am not misguided in which alternative is the morally worse one.<sup>22</sup>

---

<sup>20</sup> Reported in (Ryberg, 2004, pg. 241) and (Ryberg, 1996, pg. 169)

<sup>21</sup> However, see my observation on the *actualist bias*, on the *egoistic bias* and note 16.

<sup>22</sup> See note 16.

Huemer's error theories suggest that our judgement is clouded and not precise, not that it is wrong.<sup>23</sup> Therefore, these error theories are not sufficient to challenge QC.

RC's supporters need something more. A promising strategy to find more convincing arguments has been adopted by Torbjörn Tännsjö and Jesper Ryberg. They acknowledged the bias of underrating low-quality lives and build a whole argument out of it. Even if I do not share their view on privileged lives, explained in the next section, my strategy for defending RC is also grounded on the fact that we underrate low quality lives.

## 2.5 Tännsjö and Ryberg: living in privileged societies

Tännsjö and Ryberg believe that a life barely worth living is similar to the normal life within what we currently conceive as a privileged society.<sup>24</sup> They think that, if this is right, RC would seem far less Repugnant, that is to say, we would have strong reasons to reject QC.

Ryberg and Tännsjö believe that lives in western societies (a) do *not* have *constantly* positive wellbeing, (b) have rarely very positive wellbeing and (c) often reach a neutral wellbeing level and below. They stress that, if positive experiences outweigh negative ones in a normal life of privileged societies, this is not by a large extent.

### 2.5.1 Tännsjö's argument

Tännsjö believes that "If (...) our basic needs are satisfied, then most of us are capable of living lives that, on balance, are worth experiencing." (Tännsjö, 2004, p. 223).<sup>25</sup> However, he concludes that people in privileged societies rarely reach beyond this level because of these reasons:

1. We spend most of the time waiting for something, often in vain,
2. Often, when what we are waiting for appears, is not what we expected it to be,
3. There are good things in a privileged life, but many of them can be achieved only by suffering,
4. We often find ourselves below the line where our lives are worth living, even if momentarily.

---

<sup>23</sup> A further discussion on reliable intuitions despite biases can be found in (Pummer, 2013). Pummer focuses much on the large number bias, but with arguments very different from mine.

<sup>24</sup> Hare shares this view, too. His arguments are similar to Ryberg's and I choose to quote Ryberg's work. Hare's arguments focus on the fact that we adapt our preferences and that a privileged life does not need to deteriorate much in wellbeing for it to become neutral or worth not living, ideas that we find when examining Ryberg's work and, in a hedonistic perspective, also in Tännsjö's. Hare's ideas can be found in (Hare, 1993).

<sup>25</sup> A similar formulation of this sentence and of every other quote from (Tännsjö, 2004, pg. 223) can be found also in (Tännsjö, 1998, pg. 161-162)

Sometimes we do have a life that is more than barely worth living. It does not last long, and life does not become more than barely worth living due to material affluence, but rather

5. Life seems more than barely worth living only as “ephemeral result of such things as requited infatuation, successful creative attempts, and (...) the proper administration of drugs.” (Tännsjö, 2004, p. 223)

Tännsjö assumes hedonistic utilitarianism, that is to say, roughly, that people’s wellbeing depends on how happy they are (Tännsjö, 1998, pp. 63-95). In order to estimate how a person fares in his life, Tännsjö sums the wellbeing experienced by the individual at each moment. Given this metric, while it’s easy to understand that point 3 suggests that people’s happiness has to be adjusted according to the suffering through which it is obtained, it is unclear whether Tännsjö believes that the first two points are reason for our wellbeing to be barely positive, neutral or slightly negative. When I wait for something, or when it comes and is disappointing to some extent, is my overall wellbeing positive, negative or neutral?

But it is at least clear that according to Tännsjö, people in privileged societies overestimate the amount of wellbeing in their lives. The primary needs of people in privileged societies are satisfied; thus they spend a good amount of time on the level where life is barely worth living. However, due to the reasons listed in points 1-3, their level of wellbeing does not often go beyond that level<sup>26</sup> and sometimes a privileged person experiences negative wellbeing (for example, grief or big disappointment), as explained in point 4. Positive welfare is experienced rarely, and only for brief moments, as suggested by point 5. It seems also implied that there is a cap for how much positive wellbeing can be experienced at a certain moment by a person, otherwise positive experiences might still greatly outweigh the negative experiences. If these five points are true, positive wellbeing outweighs negative wellbeing only by a small extent. Therefore, it is not true that “a world where, on balance, each person lives a life barely worth living is a world where everyone is much worse off than are at least the happiest people in our world” (Tännsjö, 1998, p. 161).

The first objection to Tännsjö’s view is that, if people in privileged societies have lives that are barely worth living, then people in non-privileged societies have lives that are worth not living. If their life is worth not living, they should kill themselves, but they do not. Many reasons why they might not do so are listed in (Tännsjö, 2004, pg. 224): people with

---

<sup>26</sup> Perhaps points 1 and 2 are reasons for dropping their wellbeing level to the neutral level or below.

lives worth not living might have false expectations towards the future, feel obliged to others, or fail to face up to reality. Furthermore, it is plausible that evolutionary selection increased our tendency to hold on to life. Finally, recall that a life worth not living is not, according to Tännsjö, necessarily terrible, thus “having lived such a life, when death is approaching, a person may well believe that he or she has lived a life well worth living” (Tännsjö, 2004, p. 224). I find this view plausible, even if I reject the last claim because, as I show in 2.6, it implies another kind of Repugnant Conclusion.

A more challenging objection comes from (Arrhenius, 2000, pp. 48-49):<sup>27</sup>

Tännsjö is a hedonistic total utilitarian. The welfare of a life is determined by just summing the utilities of the happy and unhappy moments in life. Consider a population that consists of very short lives, say a minute of slight happiness. According to Tännsjö’s theory, these lives enjoy positive welfare. It is hard to deny that such lives have very much lower welfare than the lives led in the privileged parts of the world. Irrespective of whether there are possible lives with very high welfare, Tännsjö’s theory implies the following recasting of the Repugnant Conclusion: For any perfectly equal population with the same welfare as the people in the privileged parts of the world, there is a population of lives consisting of just one minute of slight happiness, which is better.

Tännsjö replies that he cannot imagine that a slightly happy human life can be this short; thus it is difficult to conceive how such a life would be like. He writes that “unless a human being lives for quite a long time, it is not possible for this being to develop the capacity for such experiences”, and in order to live a life worth living a human has to live both “moments that are, as such, worth not living, and other moments that are, as such, well worth living” (Tännsjö, 2004, p. 220). If a one-minute happy life is not recognizable as a human life, it would be a case of “speciesism” to claim that human lives are necessarily better than the lives of creatures that lives for a minute. In fact, we would not think that our lives are necessarily less valuable than the lives of creatures whose life span is so long that, from their perspective, our life lasts a very short time.

I am not fully convinced that humans need to live for a while in order to be capable of enjoying slight happiness. If we can experience slight happiness soon in our lives, the very

---

<sup>27</sup> In this objections lives have different durations, and in general considering cases in which lives have different durations is beyond the scope of my work. However, I still take this objection into account, because (1) it can be modified in an objection that does not consider lives with different durations, as I show below, and (2) it is relevant in order to evaluate Tännsjö’s argument.

unfortunate case of a slightly happy child whose life lasts only few days challenges Tännsjö's reply.

Furthermore, even if Tännsjö is right about child development of emotions, Arrhenius' objection can be slightly changed. Assume that a life in a coma is at the neutral level. Suppose that, instead of creatures living just one minute, we have people living in a coma for their whole life except a minute, in which they experience slight happiness. If Tännsjö's hedonistic utilitarianism and RC have to be accepted, then *Z* might look like a world in which everyone lives a minute of happiness and then a life in a coma. Tännsjö might of course bite the bullet once again, but saying that a world *Z* such as the one just described is better than a world *A* in which everyone experience a high welfare is not a sufficient defence of RC.<sup>28</sup>

### *2.5.2 Ryberg's preference satisfaction view*

Ryberg, unlike Tännsjö, adheres to a preference-satisfaction theory of wellbeing. More specifically, he thinks that a person's wellbeing depends on her "overall preference" or "global preference". The "global preference" is some sort of life plan, or basic project, that everyone has in her life: our wellbeing level is high to the extent in which our life is going according to this basic project.<sup>29</sup>

According to Ryberg, in a privileged life enough of our global preferences are satisfied and the amount of wellbeing is positive. But he claims in (Ryberg, 1996, p. 167) that satisfied preferences do not outweigh to a large extent the number of frustrated ones. Thus a privileged life is, most of the time, not far from neutrality.

However, in exploring how much a privileged life differs from neutrality, Ryberg decides in the end not to rely much on his theory of wellbeing.<sup>30</sup> I think he does so for two reasons. First, if his arguments do not hinge his theory of wellbeing, then they cannot be debunked by objecting to his theory of wellbeing. Second, I believe he thinks that, in order to understand whether or not a privileged life is barely worth living, he would still need to rely mainly on his impressions rather than the theory itself.

Thus, in comparing the life of a person living in privileged societies with a life barely worth living, Ryberg warns that

---

<sup>28</sup> I owe Krister Bykvist the discussion which sparked my two observations on Tännsjö's reply to Arrhenius.

<sup>29</sup> More on this can be found in (Ryberg, 1996, p. 167). Ryberg admits that we might have vague ideas about how to live our lives and that our global preferences might be adjusted, and sometimes radically changed, as long as our life progresses. Still, how well one's life goes depends on how much of our lives is going according to the overall preference.

<sup>30</sup> "Probably the best we can do is to present a more impressionistic account of the net well-being of a normal privileged life. In the present context I will therefore not pledge myself to any theory of well-being, but will point at a few things which I think characterize many normal privileged lives." (Ryberg, 1996, p. 166)

[His] judgment is highly impressionistic. However, as long as we do not even tentatively possess methods for measuring happiness and unhappiness in the way required here, this can hardly be avoided. This weakness, therefore, is one that encumbers any view on the matter (Ryberg, 2004, p. 242).

Ryberg thinks that, *on a daily basis*, a privileged life contains several periods of neutrality. When we are not experiencing neutrality, we experience either positive or negative welfare, but the former does not seem to greatly outweigh the latter. Thus “we would not have to remove much of what makes life worth living to turn our daily life into one at neutrality or even below” (Ryberg, 2004, p. 242).

This is claimed about the *daily* life of privileged people. When it comes to the whole life, instead, Ryberg admits the possibility that the total of wellbeing experienced in a whole life is higher in a privileged life rather than in a life barely worth living: “Once we start summing up over days and years the two lives might well differ significantly” (Ryberg, 2004, p. 243). Even if this might be true, according to Ryberg, RC remains not repugnant: if daily lives in *Z* are similar to daily lives in privileged societies, according to Ryberg, RC’s Repugnance (QC) is taken away. According to Ryberg, this would be true even if we accept the *Compounding small numbers* bias explained in 2.4, since the mistaken general impression “seems to be that a life barely worth living is significantly different also at a daily level” (Ryberg, 2004, p. 254 note 4), and Ryberg believes to have demonstrated otherwise. But not everyone agrees with Ryberg’s impressions on privileged lives being just barely worth living.

### 2.5.3 Ryberg’s and Tännsjö’s impressionistic evaluations

Ryberg is of course aware that many privileged people think that their lives could be significantly worse than they are at the moment, but still worth living. Since he explicitly relies on an impressionistic judgement, this is a problem he needs to face: if many people’s impressions are different from his, then his impression might be wrong.

I believe that by “impressionistic evaluations” Ryberg means that the evaluations are not based on clear data accessible to everyone but rather on the author’s bare intuition. Tännsjö does not support his arguments with data clearer than Ryberg’s but just with his considerations and intuitions, thus Tännsjö is as impressionistic as Ryberg and needs to answer the same objection.

Ryberg lists in (Ryberg, 2004, pp. 245-247) some possible biases that might lead people to disagree with him:

- Most people might substitute the question “do I live a life barely worth living?” with the question “would I go on with my life if my life would be worse than it actually is?”. For reasons seen discussing the suicide objection in 2.5.1, the answer to the latter would be “no” even if the answer to the former is “yes”.
- Most people might believe that, since they experience happiness often, they cannot have a life just barely worth living. Such people forget that “a normal life also contains frustrations, sorrow and many other unhappy experiences” (Ryberg, 2004, p. 245) that balance the frequent happiness. It is important to remember that our memory is highly selective, thus we might be biased towards remembering the good things in our lives rather than the bad things, perhaps due to evolutionary reasons.<sup>31</sup>

I find plausible that many people fall for the first bias, but I find hard to believe that *all* people who believe they live a life more than barely worth living are substituting the question with an easier one. Since this is so, getting rid of this bias is insufficient to believe that people living in privileged societies live lives barely worth living. The second bias analysis is not more convincing: some people might have an optimistic perspective of life and overestimate the amount of good experiences, while others might have a pessimistic perspective and overestimate the amount of suffering they live. Optimists and pessimists have different impressions and it is hard to tell who is wrong. Certainly, memory is selective, but the selection can be biased in favour of the bad moments as in favour of the good moments.

Both Ryberg and Tännsjö also extensively address a third bias. This is how Ryberg describes it in (Ryberg, 2004, p. 247):

- People adapt quickly to changes adjusting aspirations and expectations to new life conditions. This makes difficult to evaluate how other people judge their lives; for example, we might judge the life of people with handicaps or leading lives less luxurious than ours as worth not living even if they themselves have adapted and thus reliably claim to live lives not less worth living than ours. We ourselves, if victim of misfortune

---

<sup>31</sup> Notice that this argument resembles Tännsjö’s five points I listed at the beginning of 2.5.1.

become disable or start leading lives less wealthy, would adapt after a period of life worth not living and return end up being as happy as we used to be.<sup>32</sup>

Tännsjö addresses the very same problem. He considers the case of people who claim to live a life worth living despite the fact that, in the past, their life was happier. Those people, when younger, had more health, had lost less of their loved ones, felt less the weight of responsibilities and so on. It can thus be objected that there are two periods in which people experienced positive wellbeing, but in one period the wellbeing is higher than the other. This means that people in privileged societies experience at least two levels above neutrality, and at least one of them is higher than the level where life is only barely worth living. Therefore, people in privileged societies do not necessarily live a life barely worth living.

Tännsjö's answer to this objection consists mainly in doubting whether this can be true. He writes that

These people should (...) acknowledge that it is indeed very difficult to plot one's own life on a scale upon which well being is assessed. It *may* be true that, on the whole, their lives are very much worth living. And yet, (...) these people, I submit, must be somewhat reluctant when they put forward this claim. Even these people should feel that, *possibly*, they did not lead such great lives in the past. Possibly, they live now at roughly the level at which they used to live, after having lived for a while in between a life worth not living. It is indeed difficult to plot one's life securely on the level of well being assessed on the relevant scale. (Tännsjö, 2004, p. 226)

Neither Ryberg's nor Tännsjö's replies seem to me sufficient to exclude that people in contemporary societies experience more than one level of wellbeing above neutrality.

I agree with Ryberg that it is possible that a person, after a severe misfortune that makes life not worth living for a while (contracting a disability or living in severe economical struggle), adapts to it and returns to a level of wellbeing somehow similar to the one she had before the misfortune. However, this does not exclude that this person is experiencing a level of wellbeing *lower* than what she was experiencing before the misfortune, *but* both levels are positive. The level of world *Z* would then be *only* the positive level *after* the adaptation. Since the greatest majority of people living in privileged societies do not experience great

---

<sup>32</sup> Tännsjö's version of this argument can be found in (Tännsjö, 2004, pg. 225).

misfortunes in their lives, it is not true that life in privileged societies is, on average, just barely worth living.

Since Tännsjö believes that there are sub-noticeable differences between the wellbeing levels, he is even less convincing when he denies that one of the two periods of those people's lives have a wellbeing level higher than the other. It might be true that it's "difficult to plot one's life securely on the level of well being assessed on the relevant scale", but if the differences between the levels of wellbeing are imperceptible then it's hard to exclude that the period before the misfortune and the period after it have a level of wellbeing, at least imperceptibly, different. Again, since not all people in privileged societies had to adapt to misfortune, it is not true that life in privileged societies has, on average, the same wellbeing as in Z.

If there is more than one level of positive wellbeing where privileged lives are steadily found, the one similar to Z is the lowest. Accordingly, a life in Z would be not much worse than the life of a person belonging to privileged societies that has lost part of her health and loved ones. However, at this point, Tännsjö's and Ryberg's position is weakened, not discarded: privileged lives are *still* similar to some extent to lives barely worth living, and therefore it cannot be excluded that Z is not as bad as it might appear at the first impression.

To conclude, since Tännsjö's and Ryberg's theories are not based on clear data and since they do not provide sufficient reason to think that impressions different to theirs are necessarily wrong, we do not have sufficient reasons to believe that lives in privileged societies are similar to lives barely worth living.<sup>33</sup>

In general, it seems highly recommended to rely on evaluations as non-impressionistic as possible for an attack to QC. Furthermore, according to (Arrhenius, 2000, pp. 45-49), even if we were sure that lives barely worth living and privileged lives were similar, this would not make RC unproblematic unless there is a demonstration that there is no logically possible life with very high welfare.

#### 2.5.4 Sufficiently challenging QC

Arrhenius notices that, even if lives in world Z would be similar to today's privileged lives, it is still possible that there are much better lives that are weakly value-superior. In fact, what makes RC Repugnant is the *comparative* aspect expressed by QC. Thus, insofar that lives

---

<sup>33</sup> This problem does not concern Huemer, since he does not claim that privileged lives are similar to lives barely worth living.

with very high welfare are at least logically possible, Arrhenius seems to say, QC stands and the comparison between A and Z remains Repugnant.

Ryberg's and Tännsjö's counter-arguments can be summarized in their shared belief that RC's Repugnance depends merely on the apparent repugnance of world Z:

What has been considered repugnant by Parfit et al. is not the mere fact that a world with moderately but very many people, can be better than a world with not so many, but even happier persons. (...) It is the comparison with the (...) Z- world (...) that has been singled out in the discussion. What has been considered repugnant is that such a world could be better than the A-world, a less populated world with very happy people. (Tännsjö, 2004, p. 227)

They believe that, if world Z is somewhat appealing, then RC is not Repugnant, which is to say not obviously false. Since there are theorems demonstrating that a theory of Population Ethics that respects very appealing conditions and avoids RC is impossible, and since RC is not obviously false, then we should not avoid RC.

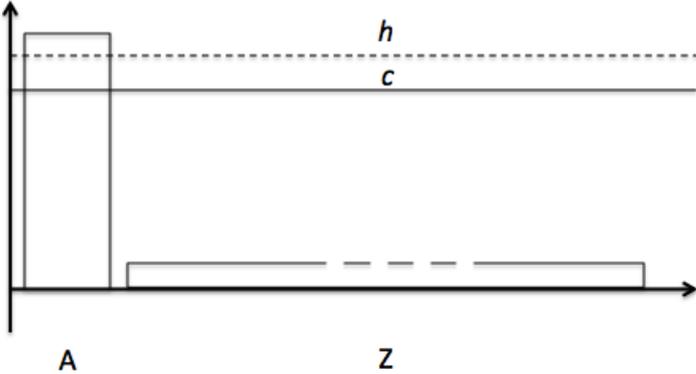
Arrhenius replies that showing that population Z is somehow appealing does not reduce our confidence in QC. Since it is the *comparison* between population A and population Z that matters, Arrhenius argues, an increase in population Z's appeal cannot challenge QC as long as lives with very high welfare are at least logically possible: population Z might be somehow good, but we can always logically imagine a weakly value-superior population A.

I argue that Arrhenius' reply is too strong. He believes that, as long as lives with very high welfare are at least logically possible, QC remains unchallenged. I argue that, if we cannot conceive how it would be like to experience a life with very high welfare which is weakly value-superior to a life barely worth living, our confidence in QC should be much diminished, regardless the logical possibility of very high welfare satisfying QC. If people in privileged societies experience a very low level of wellbeing, as Ryberg and Tännsjö believe, a very high welfare satisfying QC might be so far from our experience to be impossible to experientially imagine.

In fact, it has to be noticed that "conceiving the (logical) possibility of very high welfare" is not the same as "conceiving a life with a very high welfare level", that is, experientially imagining a very high welfare level. To explain this difference I ask the reader to describe what it would be like to live the best life he or she can possibly imagine to live. By doing so the reader has conceived a life with a very high welfare level, the highest

experientially imaginable by the reader. The reader can conceive that it is (logically) *possible* to live a life with welfare a billion times higher than the life he or she has just described: a life of a deity might be of such kind. But the reader cannot imagine how it will be like to experience a life with such a high welfare: the reader can only imagine that such a life is logically possible. There is no limit to conceiving the (logical) possibility of very high welfare, but there is a limit to what we can conceive experientially.

Consider the following graph representing populations A and Z with blocks. The rectangles' width is the population's size and their height is the welfare level experienced. The continuous line *c* is the maximum wellbeing level that we can experientially conceive and the dashed line *h* is the level of very high welfare that we think populations need to reach in order to satisfy QC.



If we cannot imagine a wellbeing level higher than *c* and if we think that the wellbeing level satisfying QC is above *h*, then QC is realized only for wellbeing levels that we cannot imagine. There are, if any, only very weak reasons to believe in a condition that can be satisfied only by wellbeing levels we cannot experientially imagine. These reasons would not be sufficient to claim that such a condition, or its variations, is a cornerstone of impossibility theorems such as Arrhenius' ones.

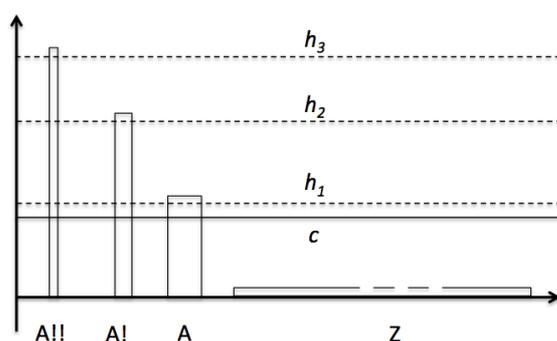
In fact, our intuitions in Population Ethics seem to require something more grounded in experience than mere logical possibility. Our evaluations concerning weak value-superiority and RC seem akin to evaluations like “torturing children for fun is morally repugnant” rather than logical evaluations like “no line has three endpoints”; in Population Ethics we are not just concerned with quantities and logical elements detached from experience, but also with wellbeing and what is morally better or worse. In order to reliably claim that a level of wellbeing is weakly value-superior to another we need to experientially imagine both of them. Logical imagination can provide no reliable intuition about wellbeing

differences concerning weak value-superiority or inferiority: it can tell us that a level of wellbeing X is high, but we need to have at least a rough idea of how it is like to live at level X to reliably state whether it is weakly value-superior to another level. If we cannot experientially imagine one level, then we have only a weak reason, if any, to believe that this level is weakly value-superior to another.

Logical imagination can, of course, be used in morality. However, as Hare points out, we generate our moral intuitions from our experience, whereas we use critical thinking, which includes logical imagination, *only* to enquire whether our intuitions can be generalized to cases that we have not experienced, thereby judging the intuition's acceptability. If some objection to a theory is raised by considering what goes way beyond our everyday experience, "no appeals to received intuitions are allowed [in the objection], because the function of critical thinking is to judge the acceptability of intuitions, and therefore it cannot without circularity invoke intuitions as premisses." (Hare, 1981, p. 131). That is to say, intuitions are applied correctly iff generated from experience and applied to logical possibilities, whereas trusting intuitions generated by merely logical possibilities is fallacious. It follows that if an objector appeals to what is "*logically possible* [...]" he has put himself beyond the range of intuition and cannot appeal to it" (Hare, 1981, p. 134). Thus QC, that is the intuition against RC, is not a convincing objection against the *Total Principle* nor can be an adequacy condition of a theory of population ethics if it works only for populations that are merely logically possible.

Notice also that, if we think that what matters in QC is just the comparative aspect between two very distant levels of wellbeing and if we admit that QC is satisfied only above what it is conceivable, we admit also that QC is reiterate *ad infinitum*: if there is a logically possible wellbeing level weakly value-superior to Z's, a wellbeing level weakly value-superior to A's is logically possible as well, and, again, a wellbeing level weakly value-superior to the wellbeing level weakly value-superior to A's is logically possible, and so on. Since every version of QC is justified in the same way, every version of QC is equally plausible. I find the existence of infinite weak value-superior levels very implausible, at least as an adequacy condition for a theory of Population Ethics.

How QC can be reiterated might be made clearer through a graph. Call A! a population whose welfare is weakly value-superior to A, A!! a population weakly value-superior to A!, call  $h_1$  the wellbeing level satisfying QC,  $h_2$  the wellbeing level after which every population is weakly value-superior than A and  $h_3$  the wellbeing level after which every population is weakly value-superior than A!. The previous graph can be expanded as follows:



The graph can be extended at pleasure with A!!!, A!!!! and so on, and with  $h_4$ ,  $h_5$  and so on.

We cannot imagine wellbeing levels higher than  $c$ , thus we have very weak reasons, if any, to state that there is at least a population whose wellbeing is above level  $h_2$  that is better than any population whose wellbeing is much below level  $h_2$  (for example A), or that there is at least a population whose wellbeing is above level  $h_3$  that is better of any population whose wellbeing is much below level  $h_3$  (for example A!). If we have little to no reason to do so, why would we have more reasons to state that there is at least a population above level  $h_1$  of wellbeing that is better of any population below level  $h_1$  (for example Z)? Why should we believe in QC if we cannot imagine a wellbeing level satisfying it?

Furthermore, if we allow intuitions to arise from comparison with logically possible lives, then QC is as plausible as *strong* value-superiority: X is strongly value-superior to Y iff any amount of X is better than any amount of Y. More specifically, it is possible to logically imagine some incredibly high welfare level X such that any number of people (including one) with welfare X is better with respect to wellbeing than any number of people with very low welfare. But notice that, if it is just the comparative aspect that matters for establishing value-superiority, that it is also possible to logically imagine some incredibly high welfare level X such that any number of people (including one) with welfare X is better with respect to wellbeing than any number of people *with the highest welfare* we can experientially imagine. If we allow intuitions to arise from comparison with merely logically possible lives, strong value-inferiority of the best lives we can experientially imagine is as plausible as QC, thus it should be an uncontroversial condition for a satisfactory theory of Population Ethics as much as QC. Strong value-inferiority of the best lives we can experientially imagine is not even nearly sufficiently intuitive to be an uncontroversial condition for a satisfactory theory of Population Ethics. Notice also that strong value-superiority can be reiterated as well: for any population X with extremely high welfare which is strongly value-superior to a set of

populations Y with lower welfare, it is logically possible to imagine a population Z with much higher welfare than X which is strongly value-superior to X.

I conclude that, if the welfare level satisfying QC is so far from our experience that we cannot even conceive it, then QC is not sufficiently justified to be an adequacy condition for a theory of Population Ethics. If the arguments of this section are correct, there is an upper and a lower cap to the wellbeing levels that can be considered for Population Ethics.

That being said, a demonstration that the welfare level satisfying QC is impossible to imagine, which would strongly suggest QC's unreliability, has not been made yet.<sup>34</sup> Furthermore, even admitting that Tännsjö and Ryberg argue successfully against QC, there is another Repugnant Conclusion to consider.

## 2.6 Reversed RC and requirements for a theory

The *Total Principle* implies:

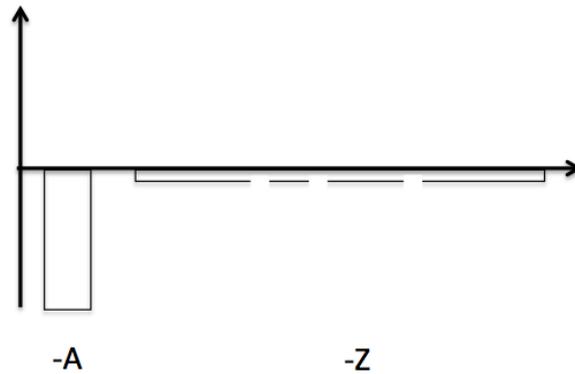
*The Reverse Repugnant Conclusion (RRC):* For any population X where each person is at a negative welfare level  $u$  (no matter how low), there is for any negative welfare level  $v > u$  (no matter how close to zero), a population Y where each person is at welfare level  $v$ , such that X is better than Y with respect to welfare. (Carlson, 1998, p. 297)<sup>35</sup>

Here is an illustration of RRC. Call  $-A$  a population with a certain number of lives, say, ten billions, at a very negative welfare and  $-Z$  a population with many more lives than  $-A$  whose wellbeing is negative, but near the neutrality level. In other words, lives in  $-Z$  are barely worth not living. The width of the rectangles is the number of people, their height is the level of welfare. The horizontal arrow is the neutrality level: positive levels of wellbeing are above it, negative levels of wellbeing are below it.

---

<sup>34</sup> Plausibly, Tännsjö and Ryberg tried to demonstrate exactly that no life with very high welfare is experientially possible. For the reasons I explained in 2.5.3, I do not think that their explanation is convincing. I try to demonstrate that no experientially imaginable life satisfies QC in 3.2. There I try to describe a life at level A using some wellbeing theories. QC will not be satisfied for any of them.

<sup>35</sup> Equivalent formulations can be found in (Mulgan, 2002, pg. 362) and (Blackorby, Bossert, Donaldson, & Fleurbaey, 1998, pg. 11).



Lives in  $-A$  are suffering enormously, lives in  $-Z$  are barely worth not living. The *Total Principle* implies that  $-Z$  is worse than  $-A$ .

Since RRC mirrors RC, RRC will feel repugnant for reasons similar to RC. Thus, presumably, RRC feels repugnant since it violates

*The Bad Quality Condition (BQC)*: There is at least one perfectly equal population with very negative welfare which is worse with respect to wellbeing than any population with slightly negative welfare.

As noticed by Tim Mulgan in (Mulgan, 2002), QC gets harder to attack as the life barely worth living gets unappealing, whereas BQC gets harder to attack as the life barely worth not living gets appealing. Mulgan implicitly assumes that life barely worth living and barely not worth living are close in the spectrum of wellbeing levels and claims that the *Total Principle* necessarily implies at least one unacceptable version of RC and RRC.

In fact, if lives barely worth living are appealing, and if lives barely worth living and barely worth not living are close in the wellbeing spectrum, then lives barely worth not living will *also* be appealing. In this case the *Total Principle* implies that a population consisting of people living, for example, in agony is better with respect to wellbeing than a much greater population of people living an appealing life. This version of RRC is unacceptable.

On the other hand, if lives barely worth not living are unappealing, and if lives barely worth living and barely worth not living are close in the wellbeing spectrum, then lives barely worth living will *also* be unappealing. In this case the *Total Principle* implies that a population consisting of people living at a very high welfare is worse with respect to wellbeing than a much greater population of people living an unappealing life. This version of RC is unacceptable.

In order to avoid the dilemma between RC and RRC, a supporter of the *Total Principle* can challenge the assumption that lives barely worth living and barely not worth living are close in the spectrum of wellbeing levels. To my knowledge, however, no supporter of the *Total Principle* has challenged this assumption. Particularly, Tännsjö seems to strongly adhere to it. As hinted in 2.5.1, he writes:

We must be careful not to think of a life worth not living as, necessarily, a terrible life. Even if there is a rather sharp line between lives worth living and lives worth not living, dropping below this line need not mean any very significant fall. A life worth not living may be a life almost worth living. And the difference between a life just worth living and a life just worth not living is trivial. (Tännsjö, 2004, p. 224)

Ryberg, too, explicitly claims that our level of wellbeing is close to neutrality.<sup>36</sup> This provides a reason to think that a life barely worth living, a life at the neutral level of wellbeing and a life barely worth not living are, in his view, close in the spectrum of wellbeing levels.

Since they claim that lives barely worth living are appealing and do not deny that such lives are close to lives barely worth not living in the wellbeing spectrum, Ryberg's and Tännsjö's arguments imply an unacceptable RRC. Huemer does not discuss these matters, thus he cannot claim that he manages to defend the *Total Principle* from RC and RRC's double threat.

The aim of the rest of this work is finding how a plausible wellbeing theory can make RC and RRC at least less problematic. Its required features seem to be:

- (a) A plausible description of life barely worth living. Such life should be appealing enough to make RC unproblematic.
- (b) A plausible description of life barely worth not living. Such life should be unappealing enough to make RRC unproblematic.

Furthermore, in order to sufficiently defend RC and RRC, this theory needs to challenge QC and BQC. As I showed at the end of 2.5.4, in order to challenge QC it is required to

---

<sup>36</sup> He claims, for example, that "we would not have to remove much of what makes life worth living to turn our daily life into one at neutrality or even below" (Ryberg, 2004, p. 242).

(c) Show that there is no experientially imaginable positive level of wellbeing that is weakly value-superior to the level where life is barely worth living.

Presumably, in order to challenge BQC it is required to

(d) Show that there is no experientially imaginable negative level of wellbeing that is weakly value-inferior to the level where a life is barely worth not living.

In section 3 I show that plausible wellbeing theories can satisfy all four conditions.

## Section 3. A gap for three wellbeing theories

### 3.1 Satisfying the requirements

#### *3.1.1 Two common features*

The theories that will be analysed in this work and that satisfy the four conditions share at least the two following features: (1) they have a clear definition of life worth living and worth not living and (2) they present a gap in wellbeing levels between a life barely worth living and a life barely worth not living. I refer to this gap as “wellbeing gap”.

Feature (1) allows me to have definitions less impressionistic than Tännsjö and Ryberg. In fact, in my work I do not identify a life barely worth living or barely worth not living by appealing to some general consideration and without data accessible to everyone. Instead, I’ll state what lives barely worth living and worth not living are by appealing to the wellbeing theories’ definitions. I thus avoid making controversial assumption on the level of wellbeing of lives in privileged societies or any other society. However, in my definitions, lives barely worth living are above the gap and lives barely worth not living below it; since the definition depends on how wide the wellbeing gap is and its width is partially impressionistic, my definitions are not completely devoid of impressionism. Anyways, as I point out in 3.1.2, I think that my impressions are widely shared.

Feature (2) means that a life barely worth living is not close to a life barely worth not living in the wellbeing spectrum.<sup>37</sup> The wellbeing gap implies the denial that the ordering of

---

<sup>37</sup> If a neutral level has to be admitted, then feature (2) states that a life barely worth living is not close to a life at the neutral level, and that a life at the neutral level is not close to a life barely worth not living in the wellbeing spectrum.

welfare levels is Dedekind complete. This is the definition of Dedekind completeness adapted to welfare levels:

An *upper bound* of [an ordered subset of successive welfare levels] is a welfare level [in the ordered set of all welfare levels] which is at least as high as every level in [the ordered subset considered]. A *least upper bound* of [said subset] is an upper bound of [the ordered subset] that is at least as low as every upper bound of [the ordered subset]. [The ordered set of all welfare levels] is Dedekind complete iff every suborder [of it] with an *upper bound* has a *least upper bound* (Carlson, Forthcoming).

Feature (2) seems necessary for satisfying requirements (a) and (b) of 2.6.<sup>38</sup> This feature is also *sufficient* to satisfy these two requirements if it is not true that:

- The worst life worth living, despite being a good life, is at a level of wellbeing low enough to be weakly value-inferior to the A level of wellbeing.
- The best life worth not living, despite being a bad life, is at a level of wellbeing high enough to be weakly value-superior to the  $-A$  level of wellbeing.

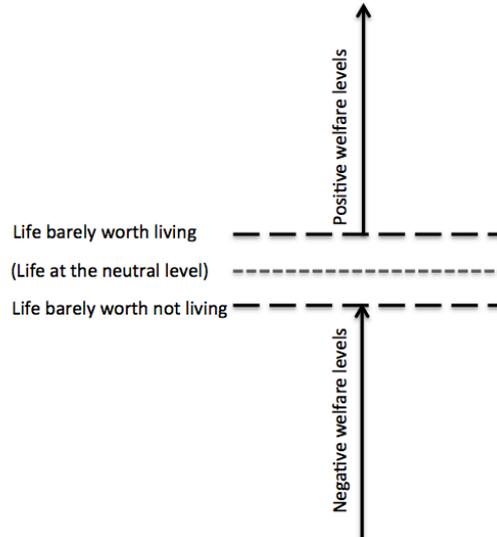
In my analysis in 3.2 I do not find any reason sufficient to believe either of them.<sup>39</sup>

I show below a graphical representation of the wellbeing gap. Each point in the vertical arrows is a level of wellbeing. The three horizontal dashed lines represent the level where life is barely worth living, at the neutral level or the level where life is barely worth not living. In order to stress that the neutral level is not necessarily present in the theories illustrated by the graph, I use brackets and the line has shorter dashes.

---

<sup>38</sup> There seem to be no common feature satisfying points (c) and (d) in the wellbeing theories I consider. It is however plausible that, for each wellbeing theory, experiential imagination has more or less the same limits in imagining the best life possible, thus (c) and (d) might be true every time (a) and (b) are true. However see note 39.

<sup>39</sup> I cannot of course conclude that these two conditions are necessarily false for all wellbeing theories adopting gaps, especially since, as I'll point out soon, I admit that the size of my gap rests on my intuition that any life worth living is much better than any life worth not living.



All theories presented in this third section have a gap between the best life worth not living (the so-called life barely worth not living) and the worst life worth living (the so-called life barely worth living). If the theory of wellbeing admits a neutral level then there are two gaps: a gap between the life barely worth living and the neutral level and a gap between the neutral level and the life barely worth not living. Notice that the wellbeing gap is compatible with both denseness and discreteness of wellbeing levels:<sup>40</sup> wellbeing levels are dense if, between a level of wellbeing and another, there are infinitely many welfare levels; wellbeing levels are discrete if, between a level of wellbeing and another, there are only finitely many welfare levels. In the case of denseness, the wellbeing gap means that positive wellbeing levels can get arbitrarily close to a lower limit, but not below. This lower limit is higher than neutrality or than any negative level. Similarly, negative wellbeing levels can get arbitrarily close to an upper limit, but not above. This upper limit is lower than neutrality or than any positive level.

The possibility of a neutral level invites an examination on what kind of lives can lay between lives worth living and worth not living in my theory. There are four possibilities:

- 1) *No neutral level.* All lives have either positive or negative wellbeing and there is not a life at a level between positive and negative. According to *the Total Principle*, all lives contribute positively or negatively to a population's value.
- 2) *Undistinguished lives.* Not all lives have either positive or negative wellbeing, and, according to *the Total Principle*, not all lives contribute positively or negatively to a population's value. There is a set of lives with an *undistinguished* wellbeing level. These lives cannot be compared to distinguished lives with respect to wellbeing and

---

<sup>40</sup> At least for the standard definition of discreteness. See (Carlson, Forthcoming) on this matter.

have some value for the population which is not positive, negative nor neutral. The wellbeing of these lives is not equally good as the wellbeing of distinguished lives, but is neither better nor worse. Undistinguished lives have been explored in (Gustafsson, Forthcoming). This possibility has the drawback of giving up full comparability between wellbeing levels.<sup>41</sup>

- 3) “*Flat*” neutral level. There is a set of lives at the neutral level of wellbeing. No life at this level is in any way better or worse than any other life at this level with respect to wellbeing and, according to *the Total Principle*, none of these lives contributes to the value of a population.
- 4) *Ordered neutral level*. There is a set of lives at the neutral level of wellbeing. According to *the Total Principle*, none of these lives contributes to the value of a population. A population *P* of any number of lives at the neutral level is worse than a population *Q* of any number of lives at any positive level and better than a population *R* of any number of lives at any negative wellbeing. However, these lives can be ordered, that is to say, some lives at the neutral level are *in some respects better than* other lives at the neutral level. This possibility has the drawback of giving up the following intuition: if a life *X* is better (worse) than a life at the neutral level *Y*, *X* has positive (negative) wellbeing.

Similarly a life barely worth living, if it has an insufficiently big decrease in wellbeing-making features, can remain neither better nor worse than it would have been without the slight decrease in wellbeing making features (“*flat*” barely positive level) or it can become *in some respects worse than* other lives at the barely positive level, despite not having a lower wellbeing level (*ordered barely positive level*). In the same way, a life barely worth not living with an insufficient increase in wellbeing-making can be a “*flat*” or *ordered barely negative level*.

To my knowledge, the possibility of ordered lives at the same wellbeing level has never been considered in the literature. However, I do not think it is so implausible. Assume hedonism and imagine a life *X* at a neutral level and a life *X1* equal to *X* in everything, except that the person living it eats one single potato chip more. If the person likes potato chips there is more pleasure in *X1* than in *X*, so *X1* is better than *X*. However, the difference in pleasure

---

<sup>41</sup> Notice that this possibility does not exclude the existence of a “*flat*” or an *ordered neutral level*.

between the lives is so trivial that, unless there is a particular story concerning that chip,<sup>42</sup> it is not absurd to claim that both lives are at the same wellbeing level. This is because, despite X1 is better than X, a person might be indifferent between living life X1 and not living it: being indifferent between living or not living a life can plausibly mean that the wellbeing level of the life considered is neutral rather than positive.

Despite making many examples to follow similar to the chip example, I will stay neutral between the possibilities just examined. For the sake of simplicity however I will assume the existence of a neutral level when considering mental state and preference satisfaction theories.

Notice that, if the gap between the best life among lives worth not living and the worst life worth living exists and is wide enough, then the word “barely” in “life barely worth living” is probably inappropriate and misleading. This means that, given wellbeing gaps, RC is avoided by the *Total Principle* for all RC formulations that define “lives barely worth living” as “lives worth living just for a trivial difference from neutrality (or from lives worth not living if there is no neutral level)”. The *Total Principle* still implies RC, instead, if “lives barely worth living” are defined as “in the set of the worst lives among lives worth living”, but I hope to show that this implication is less problematic than it seems. Henceforth I will refer to this latter interpretation of RC. Everything said in this paragraph about lives worth living and RC is also true for lives worth not living and RRC.<sup>43</sup>

### 3.1.2 Wellbeing gaps

The gap in the levels of wellbeing might seem a strange feature at first, but I hope to show that it is a reasonable assumption for at least some of the major wellbeing theories. I justify the gap mainly by claiming that what separates a life worth living from one worth not living cannot be a trivial difference in wellbeing-making features: I thus deny what Pummer called “hypersensitivity” between there being positive or negative of wellbeing levels and wellbeing-making features. Pummer defines hypersensitivity as follows: “A-properties are hypersensitive to B-properties when a slight difference with respect to B-properties makes a radical difference with respect to A-properties” (Pummer, 2019, p. 11): in my case, A-properties are the positivity, negativity or neutrality of wellbeing, and B-properties are

---

<sup>42</sup> The single chip is incredibly tasty, the person living X1 is starving or is really sad thus a single chip can make a huge difference in level of hunger or mood, chips have a symbolic meaning for the person living X1 or other very rare cases.

<sup>43</sup> The interpretation of the formulation “lives barely (not) worth living”, instead, has no impact on (B)QC.

wellbeing-making features.<sup>44</sup> I will refer to this kind of hypersensitivity as “wellbeing hypersensitivity”.

For each wellbeing theory, some greater (smaller) quantity of wellbeing-making features in a person’s life increases (decreases) this person’s wellbeing level. In other words, a wellbeing level is determined by wellbeing-making features. It seems, however, *implausible* that a very small change in wellbeing-making features has the effect of changing the wellbeing level from an absolute kind (that is positivity, negativity or neutrality) to another. In order to change the absolute kind of the wellbeing level, the improvement or worsening of wellbeing-making features must be sufficiently great. This implies that, near the limit between lives worth living and not worth living, there is a gap in the wellbeing level: if wellbeing hypersensitivity is denied, only a sufficiently large difference in wellbeing-making features can move a life from an absolute kind to another. A life’s welfare level is determined by, *and only by*, wellbeing-making features: if a small change in wellbeing-making features does not produce a change in wellbeing levels, then there is a gap in wellbeing levels.

It is unclear how wide is the wellbeing gap produced by the denial of wellbeing hypersensitivity. My intuition, which I believe is widely shared, is that the difference in wellbeing level between *any* life worth living and *any* life worth not living seems somewhat sizable: since this seems so, I assume that the wellbeing gap is sizable as well. If my definitions of life Z made from 3.2 onwards are plausible, then my assumption of a sizable wellbeing gap is plausible as well. From 3.2 onwards I will also give several examples showing why the denial of wellbeing hypersensitivity and, consequently, the existence of a gap between lives worth living and worth not living are plausible for each theory I examine.<sup>45</sup>

I am aware that denying wellbeing hypersensitivity can raise problems similar to the Sorites paradox. In fact, a small increase in wellbeing-making features, say, tasting an additional chip as in the example in pg. 32, is insufficient to change a life at the neutral level to a life at a positive level. A second small increase in wellbeing-making features, say, tasting a *second* additional chip, is also insufficient to increase the wellbeing level. But, small increase by small increase in wellbeing-making features, a sufficient amount of small increases in wellbeing-making features might eventually increase the wellbeing level. This is

---

<sup>44</sup> Pummer argues against hypersensitivity in a different context, namely in Sorites problems and spectrum arguments in Population Ethics, not in wellbeing theories.

<sup>45</sup> If the reader wants an intuitive example now, consider again the chips’ example at pg. 32/33: a person’s life cannot be worth living or not worth living depending on a single chip more or less than the neutral level. This means that a hedonistic theory, it is plausible to believe that (1) positivity, neutrality or negativity of neutral level are not hypersensitive to wellbeing-making features and (2) there is a gap between the level where life is worth living and the level where life is worth not living.

similar to the bald man paradox: losing a single hair does not make a person bald, and neither losing two does, but after a sufficient loss of hair eventually makes a man bald. I think that the Sorites problem of the wellbeing gap can be avoided by stipulating that an aggregation of nontrivial differences can hit a threshold, or have an holistic effect, and consequently become relevant for wellbeing. However I am aware that it is problematic to find where the threshold is or to describe how the holistic effect is caused, since it presumably changes from wellbeing-making feature to wellbeing-making feature. This problem is not a sufficient reason to reject my theory: it is hard to identify the threshold of the number of hairs lost above which a man is bald and below which a man is not, but we *still* have men bald and not bald.

The improvement in wellbeing-making features which is *insufficient* to increase or decrease the wellbeing level can be understood in two ways, depending on the interpretation of wellbeing levels where life is barely worth living, barely worth not living, or neutral. Consider the “*flat*” level interpretation and, say, a neutral life that trivially improves in a wellbeing-making feature.<sup>46</sup> In this case the time at which the person is experiencing an increase in a wellbeing making feature might be better, but her *life* is not better: her life is exactly as good as it would be without the small improvement. Consider instead the *ordered level* interpretation and a neutral life that insufficiently improves in a wellbeing-making feature. In this case a person’s life is slightly overall better than it would have been without the small improvement, *but* still at the neutral level.<sup>47</sup> All what said in this paragraph for a neutral level life which trivially improves is valid also for a life barely worth living which trivially worsens, a life barely worth not living which trivially improves and a neutral life which trivially worsens.

I am not the first to introduce gaps in population ethics. For example Charles Blackorby and David Donaldson<sup>48</sup> introduced critical levels. In critical level theories there is a gap between the level where life is worth living and the level where a life starts to contribute to the value of a population.

There are two main differences between a critical level theory and my theory. First, a critical level theory is a theory about how to determine the value of a population from the

---

<sup>46</sup> Of course, this example works also if the neutral life is replaced with a life barely worth living that trivially worsens, or with a life worth not living that trivially improves.

<sup>47</sup> If between the positive and the negative levels there are undistinguished lives, it is unclear in which way the improvement in wellbeing-making features which is insufficient to increase the wellbeing level should be interpreted.

<sup>48</sup> See for example (Blackorby & Donaldson, 1997).

wellbeing of the lives in a population; my theory, instead, is just about wellbeing. Second, according to a critical level theory the improvement of the life of a person to the wellbeing of a population is equal to the welfare of that person minus some positive critical level. In my proposal *all* lives with positive levels of wellbeing contribute to the wellbeing of a population. In my proposal the neutral level is not close to a positive level: there is no level of wellbeing between these two levels. This allows my proposal to avoid the classical objection to critical level theories:

*The Very Sadistic Conclusion:* For any population of lives with negative welfare, there is a population of lives with positive welfare which is worse, other things being equal. (Arrhenius, 2000, p. 73)

Consider a population A of lives worth not living and a population B much larger than A consisting in very many lives with welfare which is positive but slightly below the critical level. Since, for critical level theories, the value of *all* wellbeing levels below the critical level is negative, if population B is sufficiently large, population B is worse than population A according to critical level theories.<sup>49</sup> I avoid *the Very Sadistic Conclusion* because, in my theory, all positive wellbeing levels have positive value for the population, thus no population of lives with negative wellbeing is better than a population of lives with positive welfare.

3.1.3 *The structure of the argument*

I now outline the rest of section 3. In (Parfit, 1984, pp. 493-502) wellbeing theories are categorized in three kinds: mental state theories, preference-satisfaction theories and objective list theories. The wellbeing-making features of each theory are represented in the table.

Wellbeing theory	Wellbeing-making feature 1: <b>Degree</b>	Wellbeing-making feature 2: <b>Duration</b>	Wellbeing-making feature 3: <b>Balance</b>
------------------	---	---	--

<sup>49</sup> In (Blackorby & Donaldson, 1997, pp 216-219) is presented a version of critical level theory that avoids *the Very Sadistic Conclusion* by introducing critical ranges of wellbeing and postulating incommensurability between populations whose wellbeing are in different ranges if populations have different sizes. As Arrhenius notices in (Arrhenius, 2000, 74-78), this is very counter-intuitive. A population with low welfare is intuitively worse than, not incomparable to, a population of a different size with higher welfare.

<b>Mental state</b>	Intensity of mental state	Duration of mental state	Balance of positive and negative mental states
<b>Preference satisfaction</b>	Degree of preference	Duration of preference satisfaction	Balance of satisfied and frustrated preferences
<b>Objective list</b>	Degree to which a listed condition is satisfied	Duration of the satisfaction of a listed condition	Balance of satisfied and frustrated listed condition

For each kind of wellbeing theory I follow these steps in order to show they satisfy the requirements listed in 2.6 and attack QC:

**Step 1: (name of the theory considered).** I briefly introduce the theory.

**Step 2: *Wellbeing hypersensitivity denial.*** I show through an example why it is plausible that a small change in the theory's wellbeing-making features is *not sufficient* to change the positivity, negativity or neutrality of a life's wellbeing level. Given that hypersensitivity is denied, there is a wellbeing gap. When considering objective list theories, step 3 and step 2 are inverted.

**Step 3: *Life Z.*** I use the definitions of the theory to describe the worst life among the lives worth living: this allows me to have a non-impressionistic *definition* of a life worth living, even if I admit that the wellbeing gap's size is left to impressions.

**Step 4: *Life A.*** I describe the best life we can experientially imagine according to the wellbeing theory.

**Step 5: *Comparison.*** I compare life A and life Z, by checking whether the former seems weakly value-superior to the latter. In no theory considered there is reason to suspect that A is weakly value-superior to Z. This implies that, if the theory of wellbeing is plausible, QC cannot be an adequacy condition for a theory of Population Ethics. This means that RC is at least less problematic than currently believed.

**Step 6: *RRC.*** Since the argument against BQC perfectly mirrors the argument against QC, I present only the latter in full length. Step 6 provides the reader what is necessary to completely deduce the former.

I do not enter into the specifics, comment or defend any of the three theories that I present in what remains of this work. I do not even necessarily share them. I just want to show that there are different plausible wellbeing theories for which RC and RRC are acceptable implications of the *Total Principle*.

## 3.2 Mental state theories

### *Step 1: Mental state theories*

Mental state theories identify positive wellbeing with positive mental states, often identified with “pleasure”, and negative wellbeing with negative mental states, often identified with “pain”. Pleasure is defined either as a feeling apprehended as desirable, as for example in (Sidgwick, 1874, pg. xxviii and 131), or as a primitive, at best specified as “positive and negative hedonic tone” as in (Tännsjö, 1998, pp. 64-67). Pain is the negative of pleasure. A third possibility of a mental state theory is a pluralistic account of “pain” and “pleasure”, according to which there is not only a single axis, the pleasure-pain axis, whose balance has to be positive for a person to be happy. This possibility has not been explored much in the literature but Haybron’s description of positive and negatives mental states can be the basis for a mental state theory of this kind. According to Haybron, positive wellbeing (or happiness, according to the author’s terminology) originates from at least seven axes:

1. Peace of mind vs. anxiety.
2. Confidence vs. insecurity.
3. Uncompression vs. compression.<sup>50</sup>
4. Exuberance or vitality vs. listlessness.
5. Flow vs. boredom or ennui.
6. Joy vs. sadness.
7. Cheerfulness vs. irritability.<sup>51</sup>

Happiness arises from some function of these seven axes in which the positive axes sufficiently to outweigh the negative axes.

My argument against QC works with all three conceptions of pleasure and pain mentioned. However I refer mainly to Haybron’s axes: (R)RC is more difficult to defend when only one value out of seven values is positive (negative) rather than when the *only* relevant value is positive (negative). It should be clear to the reader that any of Haybron’s axes can be replaced with the “pleasure” - “pain” axis as defined by Tännsjö or Sidgwick.

---

<sup>50</sup> A person is “compressed” when is unable to enjoy, express and fulfill herself. For a more accurate description, see (Haybron, 2008, pg. 118-120).

<sup>51</sup> This list can be found in (Haybron, 2008, pg. 121). Also, notice that Haybron’s theory is far more complex than what briefly reported here, but this suffices for the purposes of this work.

### *Step 2: Wellbeing hypersensitivity denial*

Wellbeing-making features in mental state theories are intensity, duration and balance of positive and negative mental states. Hypersensitivity of a life's absolute value (positivity, negativity or neutrality of wellbeing) to the intensity of a mental state is implausible. Suppose that I am in an arbitrarily low but positive mental state.<sup>52</sup> Suppose then that I suddenly realize that I am running out of socks, thus I'll need to do the laundry sooner than expected and I have not booked it yet: this worries me just slightly. My "peace of mind vs anxiety" axis might change for the worse, but it is implausible that my life will become *bad* for me because of this change. The change in the axis does not seem relevant enough to become a change for the worse in my level of wellbeing. Thus, the gap is plausible for the mentioned mental state theories of wellbeing.

Similarly, a life's absolute value does not seem hypersensitive to the duration of a mental state: experiencing joy, no matter how great, for a single minute in a life will presumably not increase the wellbeing level of that life, and make it *good for me*, if I would otherwise have had a neutral life, but certainly experiencing joy for a sufficient amount of time increases a life's wellbeing level. Thus it is plausible to have a gap in amount of time for which a mental state is experienced. On what is the minimum length for the relevance of a mental state, however, intuitions differ a lot, and might change depending on the mental state considered (just for starters, negative emotions might behave differently from positive emotions).<sup>53</sup> I will not examine this matter further: I think the present analysis is sufficient to avoid objection such as Arrhenius's RC with one minute happy lives, discussed in 2.5.2, by claiming that positive or negative mental states have to be experienced for *enough* time, whatever this time would be, in order to make a life's wellbeing positive, negative or neutral.

Finally, I deny wellbeing hypersensitivity in the balance between positive and negative mental states. A life is worth living iff the balance of positive over negative mental states is *non-trivially* positive for intensity or duration. Imagine a perfect balance between positive and negative, and subtract the annoyance caused by skipping a meal, or eating less in a meal, for a couple of times: it seems implausible to claim that this life is worth living just due to this difference. Similarly, a life is worth not living iff the balance of positive over negative mental states is *non-trivially* positive for intensity or duration. A trivial predominance of positive or negative mental states, in intensity or duration, does not make a

---

<sup>52</sup> This argument works also if the starting point at positive welfare is replaced with a starting point at the neutral state.

<sup>53</sup> On this see (Carlson, Forthcoming, pp. 22-24).

life worth living rather than not living: as (Carlson, Forthcoming, pp. 22-24) argues, the difference in balance has to hit a certain threshold, or the life must be considered from an holistic point of view, in order for a life to be worth or worth not living.<sup>54</sup>

### *Step 3: Life Z*

According to the mental state theory of wellbeing so far presented, a life is barely worth living when there is a sufficiently positive difference of the axis pattern from the neutral level (or the negative level if there is no neutral level) for a sufficiently long time. Given the wellbeing gap, a positive or negative wellbeing level cannot get arbitrarily close to the neutral level.

Imagine for example a person whose joy is non-trivially positive just enough, in duration and intensity, to make her life reach the levels above the gap. Even if the level of the “joy vs sadness” axis, or of any other axis, might be even more positive, this person lives a good life. More specifically, she enjoys the lowest level of wellbeing between the positive levels. This is the level of wellbeing of Z lives.

### *Step 4: Life A*

The level of wellbeing of A lives, instead, will be the one experienced when a person’s states of mind are at the highest positive pole of the axes we can experientially imagine for the whole life.<sup>55</sup>

### *Step 5: Comparison*

In population Z, life is worth living. People’s life could be better, but they are happy. In A lives, we cannot imagine how life could be better. Regardless, I do not find obvious that, between a (relatively) big population with Z lives and a (relatively) way smaller population of people whose axes are as positive as we can experientially imagine, the latter is better with respect to wellbeing. In fact, it does not seem plausible to say that there is a weak value-superiority between, say, the joy that makes the difference between the neutral level and the positive level and the best pattern of the axes we can possibly imagine. This is supported by the following consideration: the smallest positive level of each emotion, if is able to shift the

---

<sup>54</sup> Carlson’s claim, however, is valid for an unspecified gap in wellbeing theories, not necessarily for a wellbeing gap between lives worth living and worth not living. Furthermore, the fact that there might be a balance between mental states makes difficult to defend the non-existence of neutral lives in mental state theories. In fact, Tännsjö and Sidgwick admit the existence of a neutral state (Tännsjö, 1998, pg. 70), (Sidgwick, 1874, pg. 124/125) and, even if he does not explicitly consider it, Haybron uses the neutral state in one example (Haybron, 2008, pg.138).

<sup>55</sup> Notice that this might be already beyond what we can experientially imagine. I am not sure how it would be like to experience perfect joy and perfect peace of mind at the same time. But I try to take QC at its best by considering RC’s stronger possible version.

wellbeing level from neutral to positive, is not very small and does not seem weakly value-inferior to the highest level of each emotion. For example, the joy sufficient to move us from the neutral level to the positive level does not seem weakly value-inferior to the highest joy we can experience, because the difference of the two kind of joys does not seem that big. Presumably, the aggregate of all maximally positive emotions is not weakly value-superior to the first positive level of a single emotion either. The worst life worth living is good, and the best life worth living does not seem weakly value-superior.

It is of course worse for a person to experience the “small” joy that is able to raise her to a higher level of wellbeing rather than to experience the best pattern of the axes. But (relatively) few people experiencing the best patter of axes do not seem weakly value-superior to (relatively) very many people experiencing a joy able to “move” them from the neutral state. We have therefore not sufficient reason to claim that, if a mental state theory with a wellbeing gap is plausible, QC should be an adequacy condition for Population Ethics. This means that RC would be a less implausible implication of Population Ethics.

If this is true, then our confidence in QC should be diminished, and we should be more willing to accept RC and the *Total Principle*.

#### *Step 6: RRC*

The argument for RRC and against BQC mirrors in mental state theories mirrors the argument for RC. A person might have a value of an axis, for example “peace of mind vs anxiety”, non-trivially negative just enough to make her life fall below the wellbeing gap. An enormous population whose life is bad due to anxiety does not seem weakly value-superior than a comparatively much smaller population with all axes maximally negative.

### 3.3 Preference Satisfaction Theories

#### *Step 1: Preference satisfaction theories*

According to preference satisfaction theories, a person has a positive, negative or neutral attitude, or degree of favouring, towards a certain state of affairs: the realization of a favoured state of affair increases the persons’ wellbeing, the realization of a disfavoured state of affair diminishes the person’s wellbeing and the realization of a state of affair neutral for the person does not change the wellbeing level. The higher (lower) the degree of (dis)favouring, the better (worse) the state of affair for you. The preference satisfaction theory I mainly refer to in

this chapter is Bykvist's object preferentialism as explained in (Bykvist, 2016). Bykvist's theory presents a neutral level, and I keep it in order to keep the explanations simpler.<sup>56</sup>

*Step 2: Wellbeing hypersensitivity denial*

Wellbeing-making features in preference satisfaction theories are the degrees of favouring, the duration of the satisfied preference and the balance of satisfied and dissatisfied preferences. I analyse wellbeing hypersensitivity in the order just mentioned.

Suppose that a person lives at a neutral state and has a neutral preference towards the salary she currently receives. She would prefer a higher salary to her current salary and would prefer her current salary to a lower salary. This does not necessarily mean, however, that her life would go *better* with respect of wellbeing if her salary was increased by one cent, or that her life would go *worse* if her salary were decreased by one cent, unless there is a peculiar story concerning a single cent difference. Regardless the increasing or decreasing of the salary, this person's attitude towards her life remains neutral. I think it is defensible to claim that it takes more than a trivial difference to increase her wellbeing, even if a person can have preferences between trivial differences. Since it is implausible that what distinguishes a life worth living from a life worth not living is a nontrivial difference, I conclude that there is no wellbeing hypersensitivity of a life's absolute value to the degree of preference satisfaction.

As for duration of the satisfied preference and the balance of satisfied and dissatisfied preferences, the analysis is similar to the one concerning mental state theories: imagine a preference, no matter how strong, satisfied for a minute and becoming irrelevant afterwards: this amount of time is not enough to increase the level of wellbeing of a life, for example, from neutral to positive. Similarly imagine that, in a life, the balance between preferences satisfied and dissatisfied is positive just due to a satisfaction that lasts only one minute: this does not seem enough to claim that this life is worth living rather than neutral. Thus, there is no wellbeing hypersensitivity of a life's absolute value to the duration of preference satisfaction, or to the positive balance between satisfied and dissatisfied preferences.

*Step 3: Life Z*

Assume now that a person called Iris has a set of fully informed, rational preferences towards her present condition. Since a person may have preference towards different components of her lives, Iris' set of preference is domain-specific. The following state of affair would be

---

<sup>56</sup> Note however that my argument would work regardless the existence of a neutral level.

neutral to her, and she would be indifferent between living and not living in this state of affairs:

*Iris' neutral life:* her husband is attached to her but puts just sufficient effort in the relationship; through her job she earns just enough money to survive with the husband's salary, and the job itself, despite not being boring, is not exciting or interesting either. She does not have any serious physical problem, but she is not in great shape, eats low quality food and suffers from some physical discomfort once in a while.

A life “barely” worth living, that is to say, the worst life among the lives worth living, would be, for Iris, somehow similar to the life at a neutral level but with a *significant* improvement in one domain of her life. This life would be a life in *Z* given my interpretation of Bykvist's theory of wellbeing. As an example, let us say that Iris' life “barely” worth living is as follows:

*Iris' Z life:* is similar in all aspect to *Iris' neutral life*, except that her relationship with his husband, despite still not ideal, is improved. The improvement is relevant enough for her to think that, despite living a mediocre life in all other respects and even if her life could be much better, her life is good for her. She has thus a positive attitude towards it.

Since the point of improvement in a relationship that, even if not perfect, still “moves” one's life from neutral to worth living is very subjective, I do not specify the extent of the improvement.

Someone might object that this improvement is in a domain, the relationship domain, which has obviously great influence in a person's life, and an improvement in this field would make a life better than “the worst life between the lives worth living”. If for example Iris' life were improved because she finds a hobby that she enjoys, her life would be good for her, but not as good as in the case of the improvement of the relationship. I can answer this objection. The hobby example would work the same way as the relationship example: in both cases the improvement in wellbeing-making features would cause her life to shift from neutral to worth living, so it cannot be true that in one example Iris' life is better than in the other example. Iris relationship with her husband is improved, not perfected, and the improvement is enough for her life to be good for her, but not more.

#### *Step 4: Life A*

Let us now imagine the state of affairs that, among the lives that are experientially possible to imagine, would maximally satisfy Iris' preferences. Life in A would be, for Iris, the following:

*Iris' A life:* her husband loves her very much and tries his best to spend time with her; through her job she earns a lot of money and the job itself is very rewarding for her. She is physically fit and her life is full of body pleasures, such as delicious food, great sex and entertaining sport activities. She has thus a very positive attitude towards this life.

#### *Step 5: Comparison*

Iris' A life is much better than Iris' Z life but the difference does not seem big enough to be weakly value-superior: Iris's Z life is still good, and Iris's A life does not look weakly value-superior. It seems not uncontroversial that a population consisting in (relatively) very many people whose life is as good for them as Iris's Z life is worse with respect to wellbeing than a population consisting in (relatively) very few people that are satisfied with their life as Iris is in her A life. Thus there are not sufficient reasons to claim that, if Bykvist's theory of preference satisfaction with the wellbeing gap is plausible, QC should be an adequacy condition for Population Ethics. This means that RC would be an acceptable implication of Population Ethics.

#### *Step 6: RRC*

The argument for RRC mirrors the one for RC: *Iris' -Z life* is similar in all aspect to *Iris' neutral life* except that, say, her job is frustrating enough to make her life worth not living. An enormous population of lives as bad as *Iris' -Z life* does not seem weakly value-superior than a much smaller population that has all preferences extremely dissatisfied.

### 3.4 Objective List Theory

#### *Step 1: Objective list theories*

According to objective list theories a person's wellbeing is determined by how well her life satisfies the objective conditions on the list. A person's wellbeing is positive iff the life satisfies, to a sufficient degree, sufficiently many conditions on the list. In this section I refer mainly to Martha Nussbaum's objective list since her account is influential in the literature.

Objective list theories, at least in Nussbaum's case, are concerned only with identifying what makes life worth living. Thus, in order to correctly identify the relevant wellbeing levels, I need to find some reasonable interpretation of the theory. Particularly, I need to say the extent to which each item is present in one's life in order for the life to be at a certain level of wellbeing. This is the part relevant to our purposes<sup>57</sup> of the list as it appears in (Nussbaum, 1992, p. 222):

1. Being able to have good health; to be adequately nourished; to have adequate shelter; having opportunities for sexual satisfaction; being able to move from place to place.
2. Being able to avoid unnecessary and non-beneficial pain and to have pleasurable experiences.
3. Being able to have attachments to things and persons outside ourselves; to love those who love and care for us, to grieve at their absence, in general, to love, grieve, to feel longing and gratitude.
4. Being able to form a conception of the good and to engage in critical reflection about the planning of one's own life.
5. Being able to live for and with others, to recognize and show concern for other human beings, to engage in various forms of familial and social interaction.
6. Being able to laugh, to play, to enjoy recreational activities.

*Step 3, here 2: Life Z*

The items of the list above come in degree. For example, the item "Being able to have good health; to be adequately nourished; to have adequate shelter; having opportunities for sexual satisfaction; being able to move from place to place" (Nussbaum, 1992, p. 222) can obviously be satisfied to different extents. For all items considered, the following is true: for some extent of satisfaction of this item, this item is *just sufficiently* satisfied.

If in a life *all* items that allow degrees are *just* satisfied *to a sufficient degree*, whatever that might be, and all items that do not allow degrees are satisfied, this life is barely

---

<sup>57</sup> I do not consider the items in the list that concern the length of a life and for items whose degrees of satisfaction are irrelevant, because they are not relevant to the present purpose. Items concerning the length of a life are not relevant for reasons explained in the introduction. For the other items, if they allow any degree of satisfaction at all, an improvement in those items will dubiously translate in an improvement in wellbeing and the difference between a worsening and the absence of the item is unclear. For example, the improvement from "being able to think" to "being a great thinker" seem to have only instrumental value, and it is unclear what a worsening different than "being unable to think" might be.

worth living, at least in Nussbaum's theory. This is how I interpret life Z in Nussbaum's theory. I think that my interpretation is plausible: Nussbaum writes that "a life that lacks any one of [the elements on the list], no matter what else it has, will be lacking in humanness" (Nussbaum, 1992, p. 222), or "[will be] not a good human life" (Nussbaum, 1992, p. 220). It is safe to assume that Nussbaum considers a life that does not satisfy a single element of the list as worth not living. Consequently, I assume that the neutral level does not exist according to Nussbaum's objective list theory.

*Step 2, here 3: Hypersensitivity denial*

For objective list theories, wellbeing-making features are the degree in which each item that allows degree is satisfied, the time for which it is satisfied and the balance between satisfied and dissatisfied items (in the particular case of Nussbaum's theory, a single item dissatisfied makes the whole balance negative). I analyse wellbeing hypersensitivity in the order just mentioned; the analysis is similar to the one made for the other wellbeing theories.

Subtract from the life barely worth living exactly one meal, or a fraction of it.<sup>58</sup> Unless this meal, or fraction of it, does not have a particular story, this difference seems too trivial to make one's life worth not living. Therefore, there is no wellbeing hypersensitivity of a life's absolute value (positivity, negativity or neutrality of wellbeing) to the degree in which an item of an objective list is satisfied.

Imagine that a listed item is dissatisfied to any high degree for one minute: this does not seem enough to make drop the level of wellbeing of a life from barely worth living to barely worth not living. The previous two examples imply also that, in Nussbaum's theory, if all listed items are minimally satisfied except for a single minute of dissatisfaction, or only for a very slight decrease in an item, such life is not bad enough to be considered barely worth not living.<sup>59</sup> Thus there is no wellbeing hypersensitivity of a life's absolute value to the time for which an item is satisfied and the balance between satisfied and dissatisfied items.

---

<sup>58</sup> Being adequately nourished belongs to item 1 of the list at pg. 45.

<sup>59</sup> Objective list theories less rigid than Nussbaum's, which allow that some items are not satisfied in lives with positive wellbeing, are of course possible. They would balance the degree to which some items are satisfied with the degree to which some items are dissatisfied. The gap is still plausible: the difference from a life worth living to a life not worth living cannot be that dissatisfied items outweigh satisfied items by a minute in which an item is very dissatisfied or by a slight dissatisfaction of the item.

#### *Step 4: Life A*

It is plausible to say that, in objective list theories, the best life we can possibly imagine satisfies to the highest degree all the items coming in degrees. Given Nussbaum's theory, a life in A will be similar to a life in which someone:

- 1) Is healthy, enjoys very tasty and abundant food, a greatly satisfactory sexual life, and travels in stunning places.
- 2) Has many very pleasurable experiences and no unnecessary pain.
- 3) Loves deeply as many people as it is possible by a human, and is deeply loved back.
- 4) Is fully autonomous in his life choices and has the moral awareness of the great philosophers.
- 5) Is extremely sensitive, generous and charitable and an extremely pleasurable company. Has also a great family and many meaningful friends, as implied by 3).
- 6) Has much free time for enjoy his/her life and have quality fun and laughs.<sup>60</sup>

#### *Step 5: Comparison*

A life in A where all the items are perfectly satisfied is much better than a life in Z where the items are just sufficiently satisfied. However the trade-off between few people enjoying tasty and abundant food for many more people adequately nourished does not seem repugnant, nor does the trade-off between few people having much free time and quality fun for many more people being able to enjoy recreational activities, even if to a lesser degree. By aggregating all the difference present in each item things may perhaps become more difficult to evaluate, and huge numbers of people might not help in making the comparison accurate. However, I cannot see how it is uncontroversial that a comparatively smaller population of A lives is better than a comparatively much bigger population of Z lives. Thus there are not sufficient reasons to claim that, if my interpretation of Nussbaum's theory is a plausible theory of wellbeing, QC should be an adequacy condition for Population Ethics. This means that RC would be an acceptable implication of Population Ethics.

#### *Step 6: RRC*

It is plausible to interpret a life barely worth not living in Nussbaum's theory as a life in which a person experiences all items of the list just sufficiently but one. For example, a

---

<sup>60</sup> Notice that life A, as described here, is so good that I am not sure it is experientially imaginable. I am therefore preparing to defend a very strong version of RC.

person whose life is barely worth not living has no occasion to laugh and relax:<sup>61</sup> this life is of course way better than a life in which every item that allows degrees in the objective list is maximally frustrated (the –A life in Nussbaum’s theory) but does not seem weakly value-superior at all. Thus our intuitions do not seem to justify BQC if Nussbaum’s theory is true, which implies that RRC can be accepted.

## Section 4. Conclusion

In this work I defended the Repugnant Conclusion by challenging the Quality Condition. This increases our confidence in the Total Principle in Population Ethics. I reach my conclusions by identifying a (metaphysical) wellbeing gap and a lower and upper cap for the levels of wellbeing that should be considered in Population Ethics.

I examined the existing defences of the Repugnant Conclusion. I found that Huemer did not challenge sufficiently the Quality Condition with his arguments and with the biases he found. Furthermore I found that the premises of Tännsjö and Ryberg’s solution are unsatisfactory, and I pointed out how the Total Principle as defended by these authors was too prone to the Reversed Repugnant Conclusion.

During this examination I argued that the lives that we should consider for establishing the adequacy conditions for a theory in Population Ethics are the lives that we can experientially imagine, and not merely logically possible lives. This puts an upper and lower cap to the wellbeing levels that can be considered for Population Ethics. I also argued that, to better defend the Repugnant Conclusion, what is required is a theory of wellbeing satisfying the four requirements listed in 2.6.

I suggested that, in order to satisfy the requirements, a theory of wellbeing should prescribe a gap between the welfare levels where life is worth living and the welfare levels where life is worth not living. If there is such gap, then a large population of people living the worst lives among the lives worth living might not seem weakly value-inferior to a much smaller population of people living the best life we can experientially imagine. Also, a large population of people living the best life among the lives worth not living might not seem weakly value-superior to a much smaller population of people living the worst life we can experientially imagine. I successfully verified this hypothesis with three kinds of wellbeing theories: mental state theories, preference satisfaction theories and an interpretation of

---

<sup>61</sup> If the reader does not find that a life without laughs and relax is worth not living, the reader can replace it with a life in which 5 items are similar to life in Z but she is, for example, in pain or lacking of social interaction enough that her life is bad for her.

Nussbaum's objective list theory. I justified the plausibility of the wellbeing gap for each theory.

My conclusions suggest that we should decrease our confidence in the Quality Condition and be more willing to accept both the Repugnant Conclusion and the Total Principle. The suggestion of my conclusion is strong to the extent in which (1) possible lives in Population Ethics have to be imagined as experientially possible rather than logically possible, (2) a wellbeing gap between life worth living and worth not living is plausible in wellbeing theories and (3) the wellbeing gap is wide enough. In this work there are arguments suggesting that at least the first two might be true to a large extent.

## Bibliography

Arrhenius, G. (2000). *Future Generations - A challenge for moral theory*. Uppsala, Sweden: Dissertation for the Degree of Doctor of Philosophy in Practical Philosophy presented at Uppsala University in 2000 .

Arrhenius, G. (2009). *One More Axiological Impossibility Theorem*. (E. a. Offprint from Logic, A cura di) Uppsala, Sweden: Uppsala Department of Philosophy.

Arrhenius, G. (2005). Superiority in Value. *Philosophical Studies* , 123, 97–114.

Arrhenius, G. (2017). The Impossibility of a Satisfactory Population Ethics. In E. Dzhafarov , & L. Perry, *Descriptive and Normative Approaches to Human Behavior, Advanced Series on Mathematical Psychology* (p. 1-26). World Scientific Publishing Company.

Arrhenius, G. (2003). The Very Repugnant Conclusion. In K. S. Sliwinsky, *Logic, law, morality: thirteen essays on practical philosophy in honour of Lennart Aqvist* (p. 167-180). Uppsala: Uppsala University Press.

Arrhenius, G., & Campbell, T. (2017). The Problem of Optimal Population Size. In Marc Fleurbaey, e.d., *International Panel on Social Progress 1st Annual Report*.

Blackorby, C., & Donaldson, D. (1997). Critical-Level Utilitarianism and the Population-Ethics Dilemma. *Economics and Philosophy* , 13, 197-230 .

Blackorby, C., Bossert, W., Donaldson, D., & Fleurbaey, M. (1998). Critical Levels and the (Reverse) Repugnant Conclusion. *Journal of Economics* , 67 (1), 1-15.

Bykvist, K. (2016). Preference-Based Views of Well-Being. In M. D. Fleurbaey, *The Oxford Handbook of Well-Being and Public Policy* (p. 322-347). Oxford: Oxford University Press.

Carlson, E. (2010). Generalized extensive measurement for lexicographic orders. *Journal of Mathematical Psychology* (54), 345–351.

- Carlson, E. (1998). Mere Addition and Two Trilemmas of Population Ethics . *Economics and Philosophy* , 14 (2), 283 - 306.
- Carlson, E. (Forthcoming). On Some Impossibility Theorems in Population Ethics.
- Fehige, C. (1998). A Pareto Principle for Possible People. In C. Fehige, & U. Wessels, *Preferences* (p. 508–543). Berlin: Walter de Gruyter.
- Goodrich, J. (2016). Mereological Dominance and the Logic of Better-Than. *Utilitas* , 28 (4), 361-367.
- Gustafsson, C. J. (Forthcoming). Population Axiology and the Possibility of a Fourth Category of Absolute Value. *Economics and Philosophy* .
- Hare, R. M. (1993). In R. M. Hare, *Essays on bioethics* (p. 67-83). Oxford: Clarendon Press.
- Hare, R. M. (1981). *Moral Thinking*. Oxford: Clarendon Press.
- Haybron, D. M. (2011, July 11). *Happiness*. Seen on February 18, 2019 from The Stanford Encyclopedia of Philosophy: <https://plato-stanford-edu.ezp.sub.su.se/cgi-bin/encyclopedia/archinfo.cgi?entry=happiness>
- Haybron, D. M. (2016). Mental State Approaches to Well-Being. In M. D. Fleurbaey, *The Oxford Handbook of Well-Being and Public Policy* (p. 347-377). Oxford: Oxford University Press.
- Haybron, D. M. (2008). *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford: Oxford University Press.
- Hesiod. (700 BC circa). Works and Days. *Ἔργα καὶ Ἡμέραι* .  
<http://mcllibrary.org/Hesiod/works.html>: MCL Library.
- Huemer, M. (2008). In Defence of Repugnance. *Mind* (117 (468)), 899-933.
- Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
- Leopardi, G. (1898-1900). *Zibaldone di pensieri* (Vol. 1-7). Firenze: Le Monnier.
- Mogensen, A. (forthcoming). Against large number skepticism.
- Mulgan, T. (2002). The Reverse Repugnant Conclusion. *Utilitas* , 14 (3), 360-364.
- Nussbaum, M. C. (1992). Human Functioning and Social Justice: In Defense of Aristotelian Essentialism. *Political Theory* , 20 (2), 202-246.
- Parfit, D. ((2016)). Can we avoid The Repugnant Conclusion? *Theoria* , 110-127.
- Parfit, D. (2004). Overpopulation and the Quality of Life. In & T. J. Ryberg, *The Repugnant Conclusion* (p. 7-22). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Pummer, T. (2013). Intuitions about large number cases. *Analysis* , 1 (73), 37-46.
- Pummer, T. (2019, Spring). Sorites on what matters. *currently unpublished* .

- Rachels, S. (2004). Repugnance or intransitivity: a Repugnant but forced choice. In J. Ryberg, & T. Tännsjö, *The Repugnant Conclusion: essays in Population Ethics* (p. 163-186). Dordrecht/Boston/London: Kluwer Academic Publisher.
- Ryberg, J. (1996). Is the Repugnant Conclusion Repugnant? *Philosophical Papers* (25), 161-177.
- Ryberg, J. (2004). The Repugnant Conclusion and Worthwhile Living. In T. Tännsjö, & J. Ryberg, *The Repugnant Conclusion: essays on population ethics* (p. 238-255). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Shopenhauer, A. (1859). *The World as Will and Representation*. Dover Publications; Reprint edition (1966).
- Sidgwick, H. (1874). *The Methods of Ethics* (republication of the 7th edition (1907) as published by Macmillan and Company, Limited, London;). Indianapolis: Hackett Publishing Company.
- Tännsjö, T. (1998). *Hedonistic Utilitarianism*. Edinburgh: Edinburgh University Press.
- Tännsjö, T. (2004). Why we ought to accept the Repugnant Conclusion. In T. Tännsjö, & Y. Ryberg, *The Repugnant Conclusion: essays on population ethics* (p. 219-237). Dordrecht, The Netherlands: Kluwer Academic Publishers.