

Aggregating value. Trade-offs, thresholds, and total goodness in lives

Luca Stroppa

A thesis submitted for the degree of PhD
at
FINO (University of Turin, Pavia, Vercelli, Genoa)
and the
University of St Andrews



2024

Full metadata for this item is available in
St Andrews Research Repository
at:

<https://research-repository.st-andrews.ac.uk/>

Identifier to use to cite or link to this thesis:

DOI: <https://doi.org/10.17630/sta/1097>

This item is protected by original copyright

This item is licensed under a
Creative Commons Licence

<https://creativecommons.org/licenses/by/4.0/>

Front matters

The copyright of this Dissertation rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

End User Agreement

This work is licensed under a Creative Commons Attribution-Non-Commercial-No-Derivatives 4.0 International License: <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

You are free to share, to copy, distribute and transmit the work under the following conditions:

- *Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).*
- *Non-Commercial: You may not use this work for commercial purposes.*
- *No Derivative Works - You may not alter, transform, or build upon this work, without proper citation and acknowledgement of the source.*



In case the dissertation would have found to infringe the polity of plagiarism it will be immediately expunged from the site of FINO Doctoral Consortium



UNIVERSITÀ
DEGLI STUDI
DI TORINO



UNIVERSITÀ
DEGLI STUDI
DI PAVIA

Candidate's declaration

I, Luca Stroppa, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 82,907 words in length (plus 4,000 words of appendices, excluding bibliography), has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in October 2020.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

A version of Chapter 8 is forthcoming in Synthese.

Date 30 January 2024

Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 30 January 2024

Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Luca Stroppa, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 30 January 2024

Signature of candidate

Date 30 January 2024

Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Luca Stroppa, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date

30 January 2024

Signature of candidate

Acknowledgments

As it is true for many journeys, the best part of the journey of writing this doctoral thesis have been the people I have been surrounded by. Many people have devoted me their time and intellect to improve the arguments of my thesis, and ensured that my journey was safe and joyful, despite the many academic and non-academic challenges I had to face in the completion of this work.

Among them, I am especially grateful to the members of my supervisory team. I want to thank my primary supervisor at St. Andrews, Theron Pummer, for the endless hours he spent with me discussing my ideas, replying to my emails, for calling offices to sort my boureaucratic troubles, for the many useful comments and encouraging me to explore always one argument more, to consider always an objection more, to illustrate one application more of my ideas. I want to thank my primary supervisor at FINO, Maurizio Mori, for the many hours spent at the phone with me discussing philosophical and personal problems, for motivating me to engage with many different projects, to be always more ambitious with them, and for teaching me the importance of communicating to non-philosophers the practical applications of my research. I want to thank my secondary supervisor at St. Andrews, Justin Snedegar, for his suggestions on how to navigate the academic word, for his sharp comments at my drafts and at my talks, and for teaching me how to keep a good mood in the madness of academia.

In addition to this great team of supervisors, I want to thank Matthew Clark for the many days we have spent talking about philosophy in these four years, and for the friendship we have developed. All of the most original parts of this thesis have greatly benefited from his insights and suggestions: this thesis would have been very different without Matt. And I want to thank Krister Bykvist, my Master supervisor at Stockholm University: the seeds of many ideas in this thesis can be traced back to my work with him, and he has been a continuous source of feedback and support over the years.

I am especially grateful to Emilia Wilson, invaluable office mate for the latter half of my PhD, for the never-ending intellectual engagement with my work, the support in critical moments, and the many good times spent together. I want to thank Petronella Randell for giving comments on different versions of the whole thesis multiple times(!), being willing to discuss any aspect of it, and making my time in Scotland better with her company. I want to thank Jean Gové, for the incredibly detailed feedback on my work, his wise suggestions, and the excellent meals we shared. I want to thank Lara Jost, for being an inseparable friend for all the time I lived in Crail, and for the insightful feedback. The thesis is a better work because of all of you.

Many other friends have given plenty of philosophical feedback to my work while also making my days better. I want to thank Gustav Alexandrie, Katharina Bernard, Giulia Cantamessi, Mattia Cecchinato, Joel Joseph, Tom Kaspers, Meradjuddin Oidermaa Khan, Nick Kuespert, Jann Paul Egner, Enrico Galvagni, Jacob Librizzi, David McDonald, Christopher Masterman, Colin McLean, Quentin Pharr, Sebastian Stuart Betanzos, Francisca Silva, Liam Stott, Thomas Randriamahazaka, Valérie Topf, Colin Troesken, and last but not least, Guido Tonella, who has been encouraging my philosophical growth for longer than the last ten years.

Extremely useful feedback has been given me in more professional context. During my time as a doctoral student, I have been lucky enough to gain valuable insights from Ralf Bader Jan Carter, Michael Croce, Lavinia Del Corona, Luca Fonnesu, Tomi Francis, Guy Fletcher, Eden Lin, Francesca Minerva, Alessandro Piazza, Gianfranco Pellegrino, Benjamin Sachs-Cobbe, Larry Temkin, Elliott Thornley, Mara van der Lungt, Valeria Ottonelli and Hayden Wilkinson. I furthermore want to thank audiences at the conferences and seminars at the Austin Graduate Ethics and Normativity Talk (AGENT) Conference (Austin), British Society for Ethical Theory (BSET) Graduate Conference, CEPPA Climate Conference (St. Andrews), XII ISUS Conference (Rome), Global Priority Institute (Oxford), Institute for Future Studies (Stockholm), Scienza ed etica del controllo riproduttivo conference organized by SIFES-MR, Understanding Value XI (Sheffield), and the work in progress seminars in both FINO and St. Andrews. I am especially grateful to the Workshop of Welfare in

Saarbrücken, who has discussed two chapters of my thesis in two of its editions, and its organisers Jonas Harney and Thorsten Helfer, for insisting for my presence in another edition. Special thanks at the St. Andrews Center for Ethics, Philosophy, and Public Affairs (CEPPA) and the Consulta di Bioetica ONLUS for the support and the many opportunities over the years.

Finally, many people have supported me during the hardships of the last four years, and shared with me the happiest parts of the doctorate, without necessarily discussing my research directly. Among them, I want to thank Nick Allen, Clara Fenocchi, Avery Hawkins, Jade Hawkins-Clark, Wren Hawkins-Clark, Elisa Gardino, Matteo Nizzardo, Mariam Mahmood, Reggie Mahmood, Elena Montiglio, Luca Savioli, Attilio and Caterina Strada, Maria Strada, Paola Strada, Aurelia Stroppa, Andreina Stroppa, Claudio Stroppa, Daniela Stroppa, Pierangelo Stroppa, Lia and Jimmy O'Brian, Rasmus Overmark, Patrick Winter-Larsen, the Borghebasket basketball team and the St Andrews Saints basketball team. Among my friends, I am especially grateful to Beatrice Foti and Michele Piccoli. Alice Stroppa, Enrica Strada, Francesco Stroppa, and Giuseppe Stroppa, without whom this thesis would have not been possible.

I want to add special thanks to Wlodek Rabinowicz and John Broome, who have commented extensively on this thesis in quality of pre-assessors. In particular, I thank Wlodek Rabinowicz for pressing me to better justify the problems that the Structured Range Theory has with respect to Non-Anti-Egalitarianism, which lead to the argument against Non-Anti-Egalitarianism in Chapter 7, Section 3.3.2; for encouraging me to formulate an axiological version of the Structured Range View; for many comments on transitivity of “better than”; for urging me to clarify that my theory is compatible with incommensurability, as specified at the end of Section 2.2 of Chapter 7 and much more. I thank John Broome for correcting some mistakes about the Tradeability condition, for pressing me to carefully distinguish the different senses with which I use the expression “neutral life”, especially in Chapters 5 and 6; for pressing me to formulate the Non-triviality condition in Chapter 9

in such a way that it avoids the problems highlighted by Binmore and Voorhoever; for suggesting the view I mention in Chapter 7, Section 5, and much more.

I am deeply grateful for the FINO scholarship, who has financed this research.

Table of Contents

<i>Front matters</i>	2
<i>Acknowledgments</i>	6
<i>Table of Contents</i>	10
<i>Chapter 1. Introduction: Aggregating Value</i>	13
<i>Chapter 2. Assumptions and Methods of Population Axiology</i>	23
1. Disambiguating good lives	24
2. Wellbeing theories: a very short overview	26
3. Ordering and populations	29
4. Population ethics as modeling	34
<i>Chapter 3. The Benign Addition Paradox</i>	37
1. Impossibility Theorems and the Repugnant Conclusion	38
2. The Benign Addition Paradox	41
3. Average Principle: the cost of rejecting Benign Addition	46
3.1 The Average Principle	46
3.2 Separability of Lives	47
3.3 Non-Sadism	48
4. ‘Superiority’ Theories Principle and Non-Anti-Egalitarianism	51
4.1 Value Superiority	51
4.2 Weak Superiority	54
<i>Chapter 4: Transitivity of “better than” and inferences</i>	59
1. Introduction	60
2. The case for non-transitivity	61
3. Inferential transitivity	63
4. Replies	64
4.1 The “Tie-breaker” Reply	65
4.2 Statistical Inference Reply	66
5. The Kidnapper and non-Trichotomy views	69
6. Pessimism, and its rejection	73
6.1 Known strangeness in the Pessimist ranking	75
6.2 Pessimist is incompatible with Restriction to a Range	76
<i>Chapter 5: Totalism: Between Repugnance and Greediness</i>	80
1. Totalism and Maximal Repugnance	81
2. Critical Range Views	85
3. Greediness in Broome	89
4. Maximal Greediness and the Benign Addition Paradox	93
5. Conclusion: the dilemma between Maximal Greediness and Maximal Repugnance	97
<i>Chapter 6: Barely good lives and the Intuition of Neutrality</i>	100
1. Introduction to barely good lives	101

2. Barely good lives	103
2.1 Necessary and sufficient conditions for a barely good life	103
2.2 Camus' criterion	104
2.3 Continuation Criterion.....	107
2.4 Better than neutral	110
2.5 A "test" for a good, bad, or neutral life: the Tradeability Principles	113
3. Examples of barely good life	117
3.1 Drab Life	117
3.2 Barely Conscious Lives.....	119
3.3 Rollercoaster Lives.....	120
3.4 Very Short Life	121
4. The Intuition of Neutrality	123
5. Tradeability and the Intuition of Neutrality	125
<i>Chapter 7: The Structured Range View</i>	<i>133</i>
1. Introduction.....	134
2. The Structured Range View	135
2.1 Joining a Critical Range View with the Intuition of Neutrality.....	135
2.2 Statement of the model.....	140
2.3 How to use the Structured Range View	149
3. The challenges of population ethics and the Structured Range View.....	154
3.1 Respecting Benign Addition, Non-Sadism, and Separability.....	155
3.2 Sequential Choice.....	157
3.3 The Only Acceptable Violation of Non-Anti-Egalitarianism.....	161
3.4 Transitivity	167
3.5 Out of the dilemma: Repugnance, Greediness, and the Structured Range View	172
4 Aggregating lives.....	178
4.1 The Structured Range View and Superiority views	178
4.2 The Structured Range View applied to lives.....	181
5. An alternative respecting Non-Anti-Egalitarianism	184
6. Conclusion	188
<i>Chapter 8. The Monstrous Conclusion</i>	<i>190</i>
1. Introduction	191
2. The Monstrous Conclusion.....	194
2.1 <i>The Monster</i>	194
2.2 <i>Most theories imply the Monstrous Conclusion</i>	195
3 Prioritarianism	201
3.1 <i>Asymptotic Prioritarianism</i>	201
3.2 Different Asymptotic Prioritarianism axiologies	208
4. Conclusion	212
<i>Chapter 9: Value Pluralism, Irreducibility, and Non-triviality of Values</i>	<i>215</i>
1. Introduction.....	216
2. Irreducibility and Non-Triviality	220
3. The first argument	224
4. The second argument	229
5. Hypersensitivity, Hyposensitivity, and Trumping	235

6. Conclusion	240
<i>Chapter 10. Conclusion</i>	242
<i>Appendix 1. On some attempts to distinguish Drab and Barely Conscious Lives from Very Short Lives and Rollercoaster Lives</i>	248
1. On Perfectionist goods	249
2. Imprecision in the betterness ordering.....	254
<i>Appendix 2. List of conditions and conclusions</i>	257
<i>Bibliography</i>	261

Chapter 1. Introduction: Aggregating Value

When evaluating which options is best, we often need to weigh up different aspects of the available options, and we need to understand how the values of different aspects of our options combine. In other words, we need to know how to aggregate value. Aggregating value is a crucial problem for high stakes choices both in personal lives and public policies. While aggregating value within an individual's life has been a source of dilemmas at least since the age of the ancient Greek heroes, aggregating value within a population has become particularly pressing in our century, given that we are called to answer the challenges of climate change. This thesis aims to propose a new theory to deal with theoretical puzzles arising from the literature on aggregating value, and to investigate underexplored questions in value aggregation.

A famous preamble to the Iliad consists of a difficult choice in how to aggregate value within a person's life. Just before leaving for the Trojan War, the Greek hero Achilles, king of the Myrmidons, was called to choose between two lives he could live. Achilles' mother, the goddess Thetis, foresaw Achilles' future: she revealed that, if Achilles left for Troy, he would die young but his achievements would be so glorious his fame would be immortal. If instead he stayed home, he would live a long and content but unremarkable life, and would be forgotten after death:

μήτηρ γάρ τέ μέ φησι θεὰ Θέτις ἀργυρόπεζα For my mother the goddess, silver-footed
Thetis,
διχθαδίας κῆρας φερέμεν θανάτοιο τέλος δέ. telleth me that twofold fates are bearing me
toward the doom of death.
εἰ μὲν κ' αἴθι μένων Τρώων πόλιν If I abide here and war about the city of the
ἄμφιμάχουμαι, Trojans, then lost is my home-return,
but my renown shall be imperishable;

ὄλετο μὲν μοι νόστος, ἀτὰρ κλέος ἄφθιτον

ἔσται:

εἰ δέ κεν οἴκαδ' ἴκωμι φίλην ἐς πατρίδα γαῖαν,

ὄλετό μοι κλέος ἐσθλόν, ἐπὶ δηρὸν δέ μοι αἰὼν

ἔσσεται, οὐδέ κέ μ' ὄκα τέλος θανάτοιο κιχεῖν.

(Homerus, 500, bk. IX, verses 410-416)

but if I return home to my dear native land,

lost then is my glorious renown,

yet shall my life long endure, neither shall the

doom of death come soon upon me. (Homerus,

1924)

Achilles' choice is difficult, and has incredibly high stakes: even Achilles' loved ones disagreed on what was best for him. Despite his mother doing everything in her power to make Achilles stay, Achilles ultimately chooses to leave only after being cornered by other protagonists of the Trojan War.

A core feature that makes Achilles' choice so difficult is that he is choosing whether to live a life that abounds in one of the best things in life—great achievements—but is comparatively much shorter than the alternative. If Achilles stays home, he will miss out on great achievements, but he will be in the position to enjoy his life (indeed, enjoy the life of a king) for much longer. Are the many years of unremarkable joy in Achilles' long life more valuable than the few years of outstanding achievements in his shorter life? Is a long life that is quite good at all times better than a short life that is outstanding at all times? How should we compare the value of different continuations of Achilles' life, with their differences in duration and quality at a time? To answer these questions, we need to understand how to aggregate values within a life.

Many real-life choices are similar to Achilles' first choice between a longer or shorter life. People are sometimes forced to choose between different treatments for terrible illnesses. For example, some cancer treatments (such as chemotherapy) lead patients to live longer by causing pain and significantly disrupting their everyday lives, while other treatments (such as a treatment entirely based on powerful painkillers) lead them to die sooner but with less suffering at a time. Less

tragically, choosing between different life projects can involve choosing between different lengths of life at different levels of enjoyment. For example, people may choose between a life project that is risky but intense, such as the project of being a professional mountain climber, rather than an alternative life project that is safer but more tranquil, such as being a banker. Many personal choices are of this kind. Like Achilles, in order to choose between these life-defining options, one needs to understand how to aggregate values within a life.

Questions of value aggregation, however, concern more than just individual lives. For example, Achilles has to face another difficult choice nine years into the Trojan War. This time, the choice did not only concern him, but the entire Greek army.

The Iliad begins with Achilles wishing to not continue fighting, as he was unfairly treated by the army commander, Agamemnon (Homerus, 500, book I). However, Achilles is the best Greek fighter: by refusing to fight, Achilles increases the pressure on the many Greek heroes fighting under Agamemnon. For each day Achilles stays away from the battlefield, all the other Greek warriors face a more difficult war.

This second choice Achilles is called to make involves aggregation once again, this time across *individuals* rather than across time. Is Achilles' complaint against Agamemnon more valuable than the many Greek heroes' additional struggles in war? Is a great personal loss by one or a few individuals more valuable than many smaller personal losses, spread across many individuals? How to compare the value distributions of quality of life affecting different numbers of people with different magnitudes? To answer this question, we need to understand how to aggregate values within a set of lives – that is, within a population.

Many real-life choices are also similar to Achilles' second choice between having few people (in Achilles' case, only himself) experiencing a great loss or imposing individually smaller losses, but spread across many more people. Suppose a government could build a renewable power plant, thereby making green energy available to each individual of the entire nation; but the only way to build the power plant is to relocate the few inhabitants of a village, each of them strongly connected

with the environment around the village by centuries of traditions. Or consider policies for having strict airport security, that save few people per year from terrible situations of terrorism or kidnapping, but create minor stress and annoyance to every passenger in every flight. Many other public choices are of this kind. Like Achilles, to choose between these high-stakes options, one needs to understand how to aggregate values within a population.

Real life choices between different populations, however, have a complication that Achilles never had to deal with: we can influence not only the quality of life of some members of a population, but we can influence also the *number* of people in a population.¹

For example, couples can decide whether or not to have an additional child. A family is a set of people, that is, a population: by choosing whether or not to have an additional child, the couple can control the number of people in “their” population. By the same token, we can vary the number of people in a “population” understood as a nation, as people belonging to future generations, or as a species.

A striking example where we are in the conditions to influence the number of people in the world’s population is climate change, one of the most pressing problems of our century. It is, in part, a problem of how to aggregate value when the number of existing people may vary. When it comes to fighting climate change, we may need to impose burdens on present people to benefit future people. But nations and families may also need to plan the demographic developments, in order to avoid serious problems of scarcity of resources. When making such a plan, we face two kinds of options. One option is to have fewer people, with each of them enjoying, on average, a high quality of life

¹ And the identity of these people. Problems linked to the identity of people are explored in the literature about the “non-identity problem” (Chappell, 2021, pp. 110–115; Kavka, 1982; Parfit, 1986, 2017a; E. Roberts & Wasserman, 2017; Woodward, 1986), which is a theme very close to the topic of this thesis, but which I will not explore here. It may, however, be worth having at least a footnote to explain how couples can also influence who ends up existing – that is, the identity of future people. The identity of future people is, at least partly, due to which cells were involved at conception. If two prospective parents never met, the children they would have had will never exist, and different children may exist instead. Or, if the conception of a child is delayed by, say, three months, the cells involved would be different, and the resulting individual would be different.

using the large amount of resources available. The other option is to have many more people, each of them with a lower average quality of life due to resource constraints, but having a greater total amount of quality of life across people.

Indeed, while there is extensive research about how climate change impacts the world, there is still work to do to evaluate how climate change impacts the world population. This is considered among many the elephant in the room of the discussion on the long-term consequences of climate change (Arrhenius & Andersson, 2022; Broome, 2015; Budolfson & Spears, 2021; Scovronick et al., 2017). As Broome puts it,

Changes in the population may be one of the most morally significant effects of climate change. The small chance of catastrophe may be a major component in the expected value of harm caused by climate change, and the loss of population may be a major component of the badness of catastrophe [...]. So we face a particularly intractable problem of uncertainty, which prevents us from working out what we should do. Yet we have to act; climate change will not wait while we sort ourselves out (Broome, 2012, pp. 183–185).²

To answer this question, one needs to understand how to aggregate values within a population when the number of people and their wellbeing may vary. The branch of moral philosophy that studies the value of populations is known as “population axiology”, and the related study of our obligations

² Some authors have argued, against Broome’s sense of urgency, that in the short term we do not need a conclusive theory of population ethics for policy recommendation about climate change (Arrhenius et al., 2021; Cafaro, 2022; Hedberg, 2023). A full discussion of this dispute is beyond the scope of this thesis. I limit myself to remark out that, even if it were true that we need no fine-grained theory of population ethics for current climate policy, supposedly aimed to reach zero emissions within the end of the century and therefore focused only on the next few years, we are likely need such theory for more long-term problems. And, we may need this theory soon. For example, if we were to fail to reach net zero emissions before 2050 and we have no economically accessible carbon capture technology, as seems a very real possibility, we will need to think about long-term policies aimed at co-existing with the climate catastrophe within the next 26 years. In this case, Broome’s concerns become only slightly less pressing.

towards populations is known as “population ethics” (Arrhenius, 2000b, Forthcoming; Greaves, 2017).

In this thesis, I present my research on value aggregation. It aims to provide new answers to questions about value aggregation. In particular, after examining and proposing new objections to some prominent answers to the challenges of population ethics, I propose a theory of population value that posits a “neutral range” of wellbeing levels. I call this theory the Structured Range View. The Structured Range View ranks populations by summing the total wellbeing outside this range, and then giving specific rules for which populations it is permissible to create within the range. This theory avoids important problems that similar summative theories have, known as the Repugnant Conclusion and Maximal Greediness (to be explained in Chapter 3 and 5 respectively). The Structured Range View has an implication that some have objected to, known as the violation of “Non-Anti-Egalitarianism” (to be explained in Chapter 3), but I will suggest that this violation is not as serious as it seems. Indeed, I suggest it is desirable. This theory for aggregating value within a population can be adapted to aggregate value within a life, too.

The thesis also presents two new theoretical puzzles for aggregation. I call the first “the Monstrous Conclusion”, a version of Nozick’s Utility Monster specifically for population ethics. Very few axiologies avoid the Monstrous Conclusion: I suggest that the best way to avoid the Monstrous Conclusion is an overlooked version of Prioritarianism. This version of Prioritarianism presents theoretical costs, but it is still less costly than all the alternatives to avoid the Monstrous Conclusion.

The second puzzle is a puzzle for value pluralist theories, according to which there are many dimensions of value that constitute anything that is valuable, be it a person’s wellbeing, population value, a political decision, or similar valuable matters. I argue that value pluralists are forced to choose between two of their core commitments. One commitment is that values are irreducible to one another. The other commitment is that the worth of some values is not trivial if compared to the worth of other values. For the purposes of population ethics, this simply means that value pluralists face the

same challenges as non-pluralist theories. However, this highlights a general dilemma for value pluralists, as they can maintain only one of their two core tenets.

The remainder of this “Chapter 1. Introduction: Aggregating Value” lays out the plan of the thesis. In “Chapter 2. Assumptions and Methods of Population Axiology”, I lay down some core assumptions of population axiology. One of the key assumptions will be that there is a “neutral level” of wellbeing, at which lives do not improve or worsen population value.

In “Chapter 3. The Benign Addition Paradox”, I illustrate via a famous “impossibility theorem”, known as the Benign Addition Paradox or Up and Down argument, that no theory can satisfy four very plausible adequacy conditions for a population axiology: we have to pick which adequacy condition to reject. I then proceed to analyze two attempts to reject adequacy conditions. Namely, I examine the attempts to reject the adequacy condition known as “Benign Addition”, by using the Average Theory as a token theory. Then, I analyze the attempts to reject the adequacy condition known as “Non-Anti-Egalitarianism” by appealing to the presence of “higher and lower goods”. While I find the rejection of Non-Anti-Egalitarianism plausible, I find the currently proposed theories for rejecting it unsatisfying.

In “Chapter 4: Transitivity of “better than” and inferences”, I propose a novel objection against theories that reject the adequacy condition known as “Transitivity of Better Than”. These theories cannot provide guidance in very easy choices involving partial information. I suggest this is a crucial problem for a theory. I further point out that we should reject “Pessimist” theories, according to which no life can be good, for the same incapability to deal with choices involving partial information.

In “Chapter 5: Totalism: Between Repugnance and Greediness”, I examine a dilemma discussed, among others, by Broome and Rabinowicz (Broome 2004, 2009, 2022; Rabinowicz 2009a, 2022b) concerning “Totalist” theories, that is, theories according to which the value of populations depends on the sum of individuals’ wellbeing. This dilemma is between endorsing the view that there is a single neutral level of wellbeing, where one’s life does not improve or worsen population value, or endorsing the view that there is a range of neutral levels. If there is only one neutral level, these

theories get an implausible result called “Maximal Repugnance”: for any population, there is a better population consisting of sufficiently many barely good lives, and for any population, there is a worse population consisting of lives that are just slightly worse than barely good lives. If there is a range of neutral levels, instead, we get what Broome calls the problem of “Greediness”: on “Greedy” theories, a change for the worse can be “swallowed” by a neutral change, and the conjunction of the two changes may be all things considered neutral. A particularly problematic form of the Greediness problem, which is sometimes called “Maximal Greediness”, consists actually in a violation of Benign Addition. I argue that theories with a neutral range avoid “Maximal Repugnance” in a way that is too unsatisfying to justify the violation of Benign Addition. We need a way out of the dilemma. To do so, we need to examine plausible candidates of a “barely good life” (that is, lives just above the neutral level(s)) such that, for any population, there is a better population consisting of these lives.

This examination is performed in “Chapter 6: Barely good lives”. I show that different traditions agree that a life is good iff it is better to continue it at the same average momentary wellbeing (i.e. average wellbeing at a time). On this basis, I argue that we can test whether or not a life is good, roughly, by seeing if doubling its duration and slightly decreasing its momentary wellbeing results in a better life. I call this the “Tradeability” test.

I also examine some lives that philosophers have considered successfully avoiding Maximal Repugnance. Namely, people find that “Rollercoaster Lives”, consisting of good and bad aspects that have a just barely positive balance, and “Very Short Lives”, consisting of good aspects only, but very short, are plausible candidates for barely good lives such that, for any population, there is a better population consisting of these lives. On the contrary, people find implausible that “Drab Lives”, consisting only of shallow pleasures, and “Barely Conscious Lives”, such as lives of insects, are not plausible candidates for lives such that, for any population, there is a better population consisting of these lives.

I then examine how these candidates for barely good lives fare with respect to the “Tradeability” test, which identifies whether a life is good, bad, or neutral. What I find is that

Rollercoaster Lives and Very Short Lives are good lives, while Drab Lives and Barely Conscious Lives are neutral lives. Even more surprisingly, lives that are (slightly) better or worse than Drab Lives and Barely Conscious Lives are neutral, too. This means that there is a *range* of neutral lives. This finding supports Critical Range views.

In “Chapter 7: The Structured Range View” I propose my own theory of population ethics, which I call the Structured Range View. It is a “totalist theory”, in that the value of populations depends on the sum of individuals’ wellbeing. It presents a range of neutral wellbeing levels, but it avoids both horns of the dilemmas discussed in Chapter 5. It does so by stating that a population is better than another if it has greater total wellbeing outside the range, and using what happens within the range as a tie-breaker. I then specify some (non-Greedy) rules for how to deal with ties within the range. These rules are a specification of the “Intuition of Neutrality”, an influential intuition according to which there is a range of wellbeing levels where it is not obligatory to add a life at these levels. This theory violates Non-Anti-Egalitarianism, but I argue that we should be happy with that result, and we should consider the Structured Range View among the most promising theories of population ethics.

If we aggregate lifetime wellbeing as I propose we should aggregate population value, we obtain that Drab Lives and Barely Conscious Lives are in the neutral range, while Very Short Lives and Rollercoaster Lives are above it. This reflects common intuitions.

While the Structured Range View respects many known adequacy conditions for population axiologies, there are underexplored challenges to be met in population axiology. “Chapter 8. The Monstrous Conclusion” introduces one such challenges, namely the Monstrous Conclusion, according to which, for any population, there is a better population consisting of just one individual (the Monster). The Monstrous Conclusion is deeply counterintuitive, but extremely hard to avoid. I argue that the most plausible way to avoid the Monstrous Conclusion is to endorse a version of Prioritarianism, according to which a benefit to the worse off people has greater value than the same benefit to the better off people. The specific version of Prioritarianism which can avoid the Monstrous

Conclusion has not yet received any defense, and itself has disturbing implications. While I consider the Monstrous Conclusion less disturbing than these implications, Structured Range View can (but is not required to) incorporate the required form of Prioritarianism, and thus avoid the Monstrous Conclusion.

In “Chapter 9: Value Pluralism, Irreducibility, and Non-triviality of Values” I show that problems similar to the Monstrous Conclusion generalize much beyond population axiology: whenever there are multiple dimensions of value involved, it is sometimes better to sacrifice all dimensions of value in order to improve sufficiently a single dimension. I argue that this is a problem especially for “Value Pluralism”, which is committed to giving non-trivial weight to different dimensions of value, but is also committed to avoiding a requirement that we sacrifice all dimensions of value in order to improve sufficiently a single one. To show that this is a problem for Value Pluralism in general, I will use Value Pluralism about quality of life as a case study.

Value Pluralists about quality of life are forced to choose between two principles they are deeply committed to. The first principle, Irreducibility, is that there are flourishing lives that are better than any life that is developed in only one dimension of value, such as pleasure. The second, Non-triviality, is that values have non-trivial weight relative to other values – adding a sufficient amount of one kind of value is not worse than adding a tiny amount of another value. These principles are central to Value Pluralists, but I show with an inductive argument that they are incompatible: one must be rejected. I argue that Value Pluralists should reject Irreducibility and retain Non-triviality. I show that this means accepting that, whenever there are multiple dimensions of value involved, it is sometimes better to improve one single dimension even if this means sacrificing all other dimensions of value.

I conclude by summarizing my findings and suggesting possible developments of the thesis.

Chapter 2. Assumptions and Methods of Population Axiology

Abstract. This Chapter introduces some key assumptions and terminology in population axiology. I will assume that some lives are better than some others, that there are prudentially good and bad lives, that at least in principle there is a “zero” level, and that the goodness and badness of lives can be measured on a ratio scale according to their wellbeing level. This Chapter also illustrates the methodology of the thesis, that is, I take population ethics to be an operation of modeling: in order to formulate a theory describing which population is better than which other, I consider abstract versions of populations, described only in terms of quality of life and number of people. This abstracts away from many factors, such as the distribution of fairness, virtues, or liberty in the population. This abstraction is needed in order to keep the complexity of the subject tractable. Knowing which population is better than which other still provides moral guidance at least other things being equal.

1. Disambiguating good lives

In this Chapter, I introduce some terminology and assumptions, common in population axiology, that I will mostly endorse without defending them in depth. This is not because these assumptions are beyond criticism: in philosophy, anything can be contested. Indeed, a crucial part of this thesis will consist in rejecting some assumptions concerning what I'll call the "zero level". However, to begin any inquiry, one needs to start from some initial assumptions.

First, a core topic of this thesis is *axiology*. Axiology is the study of value. In the case of population axiology, the value we are concerned with is classically goodness. Population axiology can be defined as the study of how to rank populations with respect to how good they are (Arrhenius, Forthcoming, p. 46). The main axiological problem we are concerned with in this thesis is how to aggregate value across people and within lives, and how to rank populations and lives given the best method to aggregate value.

In population axiology, we assume that some lives are *prudentially* better than others for those who live them. To see what this means, imagine a person, call them Bob, who has the aspiration of having a family and a career as a painter, but their health and economic condition frustrate these aspirations and cause this person to suffer hunger, illness, and great physical discomfort for the entirety of their life, such that not only their aspirations are frustrated, but so are any alternative aspirations Bob may have developed instead. Bob's life is *prudentially* worse than the life of a person with the same aspirations but who, due to much more favorable health and economic conditions, has a wonderful family and a successful career as a painter.

It is worth expanding on what it means for a life to be prudentially better or worse, as distinguished from other ways in which a life can be better or worse. Susan Wolf disambiguates between three ways in which lives are described as "better" or "worse" (Wolf, 1997, 2010, 2016). (1) To live a better rather than a worse life can be interpreted as to live a more rather than less *meaningful*

life - what this means exactly, if anything at all, is a matter of debate.³ (2) To live a better rather than a worse life can be interpreted as living a *morally good* life, that is, the life of a good person, rather than a less morally good life. (3) To live a better rather than a worse life has also been interpreted as meaning to live a *prudentially* better life, a life that is better *for* the person who lives it, a life with a *high quality of life* rather than a low quality of life. And of course, to live a better rather than a worse life can be seen as a combination of meaningful, morally good, or prudentially good lives.⁴

Among these, population axiologists typically refer to the *prudentially good* life when they talk about better and worse lives. There are some exceptions, such as Fred Feldman (Feldman, 1995; see Arrhenius 2000, forthcoming for a discussion) and arguably Perfectionists (See Appendix 1. On some attempts to distinguish Drab and Barely Conscious Lives from Very Short Lives and Rollercoaster Lives, Section 1). I will stick to the mainstream approach and, unless specified otherwise, when talking about some lives being better or worse than others, I will be referring to lives that are prudentially better or worse, that is, better or worse for who lives it.⁵

I refer to the degree to which a life is better or worse for who lives it as the “wellbeing level” or “welfare level”. I use the terms interchangeably. A life is (prudentially) better than another, that is, it has a higher quality of life, if it has a greater wellbeing level. A life is worse than another if it has a lower wellbeing level, and equally good if it has the same wellbeing level. I do not exclude (and indeed I often discuss the possibility) that, for some two lives, one is not better, not worse, nor equally good as the other, that is, they may be incommensurable (Andersson & Herlitz, 2022; Chang, 1997, 2002, 2015a; Rabinowicz, 2009a, 2022b; Raz, 1986). The same may be true for some two momentary wellbeing levels.

³ See (Camus & O’Brien, 1991; Nagel, 1971; Sartre, 1973) for some classic discussion on the topic. For more recent discussion, see (Metz, 2002; Mori, 2014; Wolf, 1997, 2010, 2016).

⁴ It is not always easy to identify what interpretation of the question some authors refer to, as they are not always careful in distinguishing these terms (Fumagalli, 2018).

⁵ It may be worth noting at least in a footnote that a given life is equally good for whoever might live it.

One of the core tasks of a population axiology is to establish in what way each life increases population value.⁶ If the lives in a population have higher rather than lower wellbeing, it is better. There are at least two broadly agreed upon features of a life that can increase population value. The first is the *intrinsic* value of a life, which is typically taken to correlate with the wellbeing of a life, as we saw above. The second is the *instrumental* value of a life. There may be some lives that increase population value independently from their own wellbeing – for example, because they are an altruistic person’s life and increase someone else’s wellbeing. Or, they may decrease population value by decreasing other people’s wellbeing. The instrumental value of lives is the impact of lives on a population that happens independently from their wellbeing.

Since wellbeing is a crucial tool for our analysis, it may be helpful to briefly illustrate what are the most influential different ways to think about wellbeing.

2. Wellbeing theories: a very short overview

In this thesis I will not take an explicit position as to what makes lives better or worse – that is, on what constitutes wellbeing. However, it may be important to point out that three main families of theories are identified as candidates for what constitutes wellbeing.⁷

One family is *Hedonism*, according to which the quality of life depends on the balance of pleasure over pain. Pleasure is defined either as a feeling apprehended as desirable, as for example in (Sidgwick, 1981, pp. 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. Some theorists admit the presence of “higher” pleasure, that are worth more than any amount of “lower” pleasure: we will discuss higher and lower goods in Section 4 of Chapter 3. The Benign Addition Paradox . For an

⁶ Arrhenius calls it the “contributive value” of lives (2000 p.7, forthcoming p.9)

⁷ I am using Parfit’s classification in (Parfit, 1986, pp. 493–502), that has become canonical. There are, of course, difficulties with this classification, as some theories seem to belong to more than one category, and some to no category at all: for suggestions of alternative classifications, see for example (Fletcher, 2013).

overview of Hedonism, see (Haybron, 2016). For some recent discussions of Hedonism, see (Bramble, 2016; Crisp, 2006; Tännsjö, 1998).

Another family is known as *Objective List Theories*. According to these theories, there is a list of wellbeing components that cannot be reduced to one single component. While pleasure typically appears on the list of what makes life better, it is joined by other items, such as meaningful relationships, achievements, or virtue. There is disagreement as to which items are on the list, and as to which criterion should be used to identify which items are to be included in the list. However, Objective List Theories agree that the better lives are the ones with more of the good things and less of the bad things. For an overview of Objective List Theories, see (Hurka, 2016). For some recent discussions of Objective List Theories, see (Bradford, 2015; Fletcher, 2013; Rice, 2013). Chapter 9 of this thesis presents a novel objection against a good number of Objective List Theories. Perfectionism, mentioned in Appendix 1, is an Objective List Theory.

The last family of wellbeing theories are *Desire-based Theories*, according to which one's wellbeing entirely depends on one's desires or preferences. These theories are commonly defined as theories according to which:

x is better for you than y iff (if and only if) you prefer x to y,

x is equally as good for you as y iff you are indifferent between x and y. (Bykvist 2016 p. 3;

Broome 2003 p. 9, Gibbard 1992 p. 68)

Where x and y are “either alternative social states, alternative lives of yours, or alternative consumption bundles” (Bykvist p. 3). A life is better the more the preferences in this life are satisfied, and the stronger the satisfied preferences are.⁸ A life is worse the more one's preferences are

⁸ I am purposefully setting aside important difficulties and disagreements in how to describe these accounts, such as different orders of preferences, conflicting preferences, whether what matters is that we obtain the object of our

frustrated, and the higher the degree of con-attitude towards the frustration of these preferences – but how to aggregate lives whose preferences are satisfied in different numbers and different degrees is an open question directly related to the main topic of the thesis. For an overview of Desire-Based Theories, see (Bykvist, 2016). For some recent discussions of Desire-Based Theories, see (Arrhenius, 2005; Arrhenius & Rabinowicz, 2015; Griffin, 2002, p. 20; Jensen, 2008; Nebel, 2022, pp. 86–89). See also (Dorsey, 2021; Heathwood, 2019; Lin, 2022).

In this thesis, I commit to no specific wellbeing theory, in order to ensure that the conclusions of my thesis do not depend on assumptions concerning wellbeing theories. I will often use examples that seem to assume one of these theories, typically Objective List Theories. I am, however, assuming no such theory: any background theory is incidental and the reader can substitute the example for one that fits with their preferred theory of wellbeing. While wellbeing theories disagree as to *why* something is good, I do not need to commit to any such explanation, as they largely agree as to *what things* are good. To see the extent to which different theories agree on goods, consider meaningful relationships. Objective List Theories say that meaningful relationships are *intrinsically* good, hedonist theories say that meaningful relationships are good because they generate pleasure in those who participate in them (*instrumental* good), and Desire Theories say that meaningful relationships are good because people desire them.

It may, however, be worth stressing how to read some of the research questions appearing in this thesis if one believes in a desire theory. Consider again Achilles, forced to choose between a short life of great achievements or a long life of unremarkable pleasure. For Objective List Theories and Hedonist theories, this choice represents a dilemma between having the greatest average of momentary wellbeing rather than the greatest total of momentary wellbeing.⁹ If one is a desire theorist, one may be tempted to conclude that Achilles should just introspect more to identify whether

satisfaction or the compound of state of affairs “us preferring something and something obtaining”, and more. See (Bykvist, 2016) for these difficulties.

⁹ Objective List Theories may have the additional dilemma of what item of the list to realise.

he prefers a shorter life of achievements or a longer life of pleasure. This is the case only on theories according to which the relevant preference for deciding one's wellbeing is the preference concerning one's whole life. On other desire based theories, this is not the case: Achilles has to decide whether to satisfy a smaller number of preferences, that he prefers more strongly, or to satisfy a greater number of preferences that he prefers less strongly – that is, he is deciding between a greater average preference satisfaction or a greater total preference satisfaction (Heathwood, 2020). The question of how to aggregate the good in each of his possible lives is thus crucial to Achilles's decision.

3. Ordering and populations

The assumption that at least some lives can be better, worse, and as good as one another means accepting another assumption: that at least some lives can be ranked. Given what I have assumed so far, we are justified in endorsing only an *ordinal* ranking, that is, we can specify which lives are better than, worse than, or as good as which other, and nothing more. But some population axiologists endorse something stronger than that about the ordering of lives, and so will I in this thesis.

I assume that we can meaningfully speak of lives being *much* better or much worse than others. To see why I am justified in thinking that there is a magnitude in what is sometimes called the “ranking of betterness”, consider again a person whose aspirations of having, say, a family and a career as a painter are fulfilled, and a person whose aspirations are frustrated. I have assumed that we can justifiably say that the first life is better than the second life. It is very compelling to think that *the extent to which* the first life is better than the second depends on the degree to which the first person's aspirations are fulfilled, and on the degree in which the second person's aspirations are frustrated. One's aspiration to have a career as a painter is satisfied to a greater degree if one is a world-famous artist rather than if one simply makes a living out of painting, and one's aspirations to have a family are frustrated to a greater degree if one is actively detested by everyone surrounding them rather than if one simply is unlucky in love.

Since, for example, projects can be fulfilled to a greater or a lesser degree, and lives may be better or worse, for example, correspondingly to how good or bad one's life aspirations are fulfilled, we can meaningfully talk about some lives being better than some others to a greater rather than a smaller extent. That is to say, population ethicists assume not only that we can meaningfully say that some lives are better than others, but also that we can meaningfully say that there can be degrees as to how much a life can be better than another. This means that we can rank lives not only *ordinally*, that is specifying only that some are better and some worse, but also on an *interval scale*, that is, there can be a difference in the extent to which some goods are better or worse than some others.

Some population ethicists also endorse what follows. They hold not only that lives can be better or worse, but also that lives can be prudentially *good* or *bad*, in absolute, non-comparative terms, for those who live them. For example, a life that consists exclusively in pure bliss is a good life, in absolute, non-comparative terms, and a life that consists exclusively in excruciating torture is a bad life in absolute, non-comparative terms (See (Mori, 2013, pp. 336–356)). There are influential axiological traditions that have denied the existence of good lives, and some have denied the existence of bad lives: I argued against them in (Stroppa, 2021), and I will offer new arguments against some of them in Chapter 4. Except when discussing these views, I assume that lives can be good or bad in absolute, non-comparative terms.

An important feature of the betterness ordering of lives is that there is a smooth spectrum that goes from any bad life to any good life. This means that one can transition from any good life to any bad life in a finite sequence of arbitrarily small steps, so long as the steps have all the same size. More precisely, most population axiologists assume that there is a Smooth Continuum of welfare levels (roughly, wellbeing levels can be represented by rational numbers). The smooth continuum implies that, for any two welfare levels, and for any number of steps, there is some sufficiently small welfare difference size of welfare difference such that there is a sequence consisting of that number of steps between any two welfare levels. Arrhenius offers the following argument to explain the intuitive force of Smooth Continuum.

One can get from one welfare level to another in a finite number of steps of intuitively slight welfare difference. Examples of such welfare differences could be some minor pain or pleasure or a shortening of life with a minute or two. These differences do not have to be of the same size or type. Let's say that a life of type *a* has higher welfare than a life of type *b*, and suppose that you are successively making *a* slightly worse, perhaps by shortening it by a minute or two or by adding some minor pain. [Smooth continuum] implies that there is a finite (but possibly great) number of such slight worsenings from *a* to another type of life *c* such that a life of this type will have the same welfare or lower welfare than a life of type *b*. It is quite hard to deny the intuitive force of this assumption. (Arrhenius, 2016, p.171)

Many population ethicists assume that, between good and bad lives, there is a wellbeing level where lives are neither good *nor* bad. That is, the assumption is that, on any wellbeing theory one may adopt, the scale for measuring this wellbeing has a "neutral level", or "zero level", where lives are neither good nor bad. I will endorse this assumption, and build on it in Chapter 7. In Chapter 5, and only in Chapter 5, I will endorse a slightly different definition of "neutral level", defined as the level where adding a life does not make a population better or worse. In Chapter 6 I will briefly justify the assumption that the two definitions actually refer to the same wellbeing level(s).

The suggestion that the existence of good and bad lives implies a "neutral level" can be traced back at least to Henry Sidgwick. Sidgwick can be read as a Hedonist about wellbeing: he thought that a life was good if there was more net pleasure over pain, and a life was bad if there was more net pain

over pleasure.¹⁰ This lead him “to the assumption of a hedonistic zero, or perfectly neutral feeling, as a point from which the positive quantity of pleasures may be measured.” (Sidgwick, 1981, p. 94).¹¹

This intuitive definition of a “neutral level” is sufficient to have a *ratio scale* of lives. In a cardinal ranking, we can not only say that some lives are better than others by some quantity, but we can also say that there is an absolute sense in which lives are good or bad. We thus assume, for simplicity, that, at least in principle, we can quantify the goodness and badness of lives (setting aside the question of how exactly we might do this). This (admittedly simplistic) quantification is a standard tool for comparing populations.

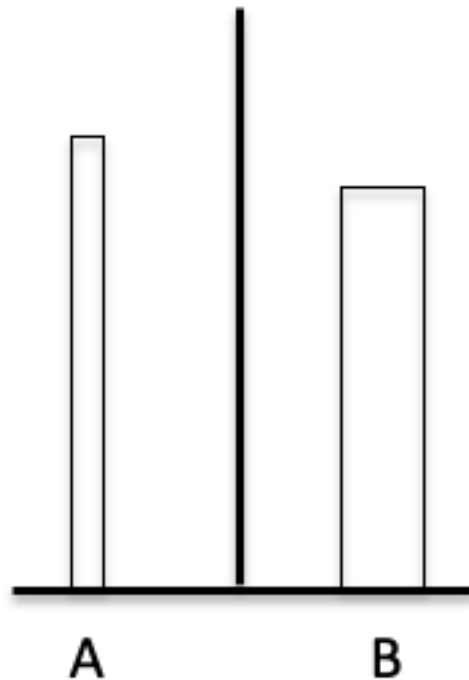
Many population axiologists represent lives’ wellbeing with rational numbers, or with lines in a graph. I too will use graphs and numbers, either because I am reporting arguments that are classically joined by graphs or real numbers (as will happen in Chapter 3), or because I worry that my text is too obscure or cumbersome without a graph or number. When I am using these graphs and numbers, I am representing wellbeing as comparable with the same precision of the cardinal scale of real numbers (unless specified otherwise). This is a modelling assumption, an idealization, as it is likely that the comparisons in wellbeing across people is not as precise as a representation with numbers would suggest. However, it is an idealization with great representational power, and crucially, not much seems to be lost by assuming this kind of precision (See (Broome, 2022)).

To see how a graph of population ethics will look like, I represent the first populations to ever show up in a graph in *Reasons and Persons* (1986), A and B, from page 385. Each square represents a population, labeled with a capital letter. The width of the square is the number of people, and the height is their wellbeing level. Occasionally, I will specify the number of people and I will represent

¹⁰ Sidgwick can also be read as a desire-based theorist, since he defines pains and pleasures in terms of preferences in favour and preferences against some mental states (1981, p. xxviii, 125-131).

¹¹ We will discuss at length how to best characterize the “zero level”, and several of the definitions that have been proposed, in “Chapter 6: Barely good lives”. For now, by “neutral level” and “zero level” I will simply mean the wellbeing level that separates good and bad lives, and at which it is neither good nor bad for a person to live.

the wellbeing level with a real number. The thicker line at the bottom is the “zero” level, and there is a thick line dividing the two populations.



When I refer to a *population* in this text, I refer to a set of lives. When doing population axiology, one tries to capture the value of *any* logically possible population (Arrhenius, 2000b, pp. 31–35, Forthcoming, p. 41; Beckstead & Thomas, 2023; Gustafsson, 2020; Huemer, 2008; Thomas, 2018). This means that, unless specified otherwise, I will not impose any constraints on whether I am discussing animal or human individuals, past, present, or future populations, and so on. I will refer to populations that have an outlandish amount of resources, and to populations where individuals have it much better than how contemporary humans have it. While we should put less credence on our intuitions involving populations far removed from the population of the actual world, sometimes these cases represent crucial problems for some theories, and we cannot simply ignore them.

This concludes the list of the assumptions that, while controversial in some debates, are widely shared in population ethics. However, to fully evaluate the text, it is important to say a few words on methodology.

4. Population ethics as modeling

Some may think that restricting the examination of population value to only the number of people and their welfare makes population ethics extremely limited in scope. Specifically, some may suspect that considering only the number of people and their welfare makes population axiology a tool only for welfarist consequentialists, who think that all that matters morally is how our choices impact people's wellbeing.

However, the scope of population ethics is broader than that, as it is relevant for any ethical theory according to which people's wellbeing has weight for practical reasoning, at least other things being equal. If one is a welfarist consequentialist, then considerations other than wellbeing are not intrinsically relevant and population ethics gives practical guidance in a quite comprehensive number of cases. However, that people's wellbeing gives practical guidance at least *ceteris paribus* is an overwhelmingly common assumption among non-consequentialists. As Rawls puts it:

Deontological theories are defined as non-teleological ones, not as views that characterize the rightness of institutions and acts independently from their consequences. All ethical doctrines worth our attention take consequences into account in judging rightness. One which did not would simply be irrational, crazy. (Rawls, 2003, p. 26)

This means that the results of population ethics are relevant for any theory according to which considerations of goodness count, at least other things being equal, and that we can test our intuitions about the betterness ordering of populations by appealing to ethical considerations.¹²

Some may still complain that, if population ethics has to be of relevance for non-consequentialist theories, then it needs to expand the number of variables involved in the analysis. In fact, if we want to leave it open whether non-consequentialist considerations, such as justice or virtues, may trump considerations about wellbeing, one may want to incorporate these considerations in the analysis.

However, population axiology is very complex even if we consider few variables and, to date, the attempts to improve population axiology by adding more variables have had particularly bad results (Arrhenius, 2000b, pp. 104–111, 139–150, Forthcoming, pp. 161–197; Feldman, 1995): it seems more promising to try to work out how to behave when considering few variables before adding other variables. Population ethics has to be considered as a *model*, where we are examining what happens with very few variables, and further relevant variables have to be added in at a later stage of research, if needed (Roussos, 2021, 2022).¹³ The examination of which further variables may be relevant is a discussion of normative theory rather than population ethics. Since I am assuming that goodness is the only morally relevant variable discussed in this thesis, I will assume what Rawls calls “teleology”, that is, that “the good is defined independently from the right, and then the right is defined as that which maximizes the good” (Rawls, 1972, p. 24).¹⁴ Non-consequentialist readers are welcome to abandon teleology as soon as morally relevant variables other than goodness are involved.

¹² Unless we think that other things can never be equal, or that the goodness of a population can never be separated from other morally relevant aspects. I know of no one arguing for this, so I will not discuss it.

¹³ I will provide reasons in Chapter 9: Value Pluralism, Irreducibility, and Non-triviality of Values to believe that population axiology would not change greatly if we considered more variables.

¹⁴ For a discussion on teleology, see also (Broome, 1991, Chapter 1, 1992).

To summarize this Chapter, in this thesis I will assume that some lives are better than some others, that there are good and bad lives, that there is a “zero” level, and that the goodness and badness of lives can be measured on a cardinal scale according to their wellbeing level. I conceive population axiology as a model, that tells us which population is best when the number of people and their wellbeing may vary. This gives practical guidance, other things being equal. As to what happens when other things are not equal, that’s a matter for research that goes beyond the scope of this thesis. We now have all we need to start a critical examination of the literature on population axiology.

Chapter 3. The Benign Addition Paradox

Abstract. This Chapter aims to provide the reader with the tools necessary to evaluate a successful population axiology. Namely, in this Chapter, I introduce an “impossibility theorem” in population axiology: no theory of population axiology can respect some set of adequacy conditions, known as Benign Addition, Non-Anti-Egalitarianism, Transitivity of “better than”, and the avoidance of the Repugnant Conclusion. Thus, all theories in population ethics violate at least one condition. In order to be fully convincing, each theory needs to provide some justification for why violating a specific condition is acceptable. In this Chapter, I report the most known arguments highlighting why rejecting the adequacy condition known as “Benign Addition” is particularly unpromising. I also consider the arguments for rejecting the adequacy condition known as “Non-Anti-Egalitarianism”. Since I will end up rejecting “Non-Anti-Egalitarianism”, it deserves attention.

1. Impossibility Theorems and the Repugnant Conclusion

Putting fresh bedsheets on a bed can be extremely satisfying. If the bedsheets are the right measure, after the fourth corner of the bedsheets has been tucked in, the result will be a smoothly, tightly packaged bed, that calls for comfortable sleep.

Putting fresh bedsheets on a bed can also be extremely frustrating. If the bedsheets are too short, once the third corner of the bedsheets has been tucked in, there is no way to insert the fourth one without one of the previous corners detaching from the bed. No matter how many times one tries, the four corners can never be all simultaneously tucked in, and if one has no other bedsheets, one is simply left with the annoying task of picking which corner is less crucial to tuck in for sleeping well.

Population axiology is a bedsheets too small for its bed. We want theories of population axiology to avoid some *prima facie* very disturbing implications. However, there are over twenty formal arguments showing that any population axiology will necessarily have some disturbing implication (see footnote 15 for bibliographic information about some remarkable arguments of this kind). We will see one such argument in greater detail in the next Section. The work of population axiologists consists mostly in finding which very disturbing implication is not that disturbing after all – much like we try to find out which corner is less crucial to cover for sleeping well when our coversheets are too short.

These arguments showing that any population axiology will necessarily have some disturbing implications are known as “impossibility theorems”. The requirement of avoiding a specific

¹⁵ The following papers present at least a theorem with completely different adequacy conditions from the others: (Carlson, 1998; Cowen, 1996; Kitcher, 2000; Nebel, 2019; Ng, 1989). Arrhenius presents six groundbreaking theorems in his (Arrhenius, 2000b, Forthcoming); the 6th theorem is considered a paradigmatically strong result. See (Carlson, 2022; Thomas, 2018, 2022a) for important comments to these theorems. The first three impossibility theorems in the history of population ethics can be found in (Parfit 1986, 419-442). Some other notable theorems are: a demonstration that all notable population axiologies (including averagism) imply the Repugnant Conclusion (Spears & Budolfson, 2019, 2021), theorems without transitivity of “better than” (Arrhenius, 2022) and theorems strengthened by considerations about probability (Arrhenius & Stefánsson, 2023; Beckstead & Thomas, 2023; Thornley, 2021).

disturbing conclusion is often called a “desideratum” for population axiology, or an “adequacy condition” for population axiology. This terminology has become standard after some impossibility theorems in social choice theory, most notably the theorem in aggregations of voter’s rankings in (Arrow, 1950).

While adequacy conditions vary in impossibility theorems for a satisfactory population axiology, one remains constant. It is the avoidance of:

The Repugnant Conclusion: For each population A of very many people— say, ten billion— all of whom have a very high quality of life, there is a better population Z consisting of some much larger number of people who would have lives that are barely good. (Adapted from Parfit, 2004, p.10)

As the name suggests, many find the *Repugnant Conclusion* unpalatable – indeed, repugnant. It is hard to believe that ten billion awesome lives are worse than sufficiently many barely good lives. However, the impossibility theorems show that the Repugnant Conclusion can only be avoided at the price of accepting alternative unappealing implications. As we will see in Chapter 5, some consider the impossibility theorems simply arguments in favour of the Repugnant Conclusion.¹⁶

To make things even more worrying, many impossibility theorems for satisfactory population axiologies can be replicated as analogous impossibility theorems for satisfactory theories aggregating momentary wellbeing (Arrhenius, 2005; Cowen, 1996; Heathwood, 2020; McTaggart & Broad, 1968; Nebel, 2022; Parfit, 1986, pp. 493– 502; Rachels, 2004). The intra-personal parallel of the Repugnant Conclusion, that these theories have been trying to avoid, is the following.

¹⁶ Sometimes an argument can be interpreted as an impossibility theorem with one desideratum presented as much less plausible than others.

Suppose that I can choose between two futures. I could live for another 100 years, all of an extremely high quality. Call this the Century of Ecstasy. I could instead live for ever, with a life that would always be barely worth living. Though there would be nothing bad in this life, the only good things would be muzak and potatoes. Call this the Drab Eternity.

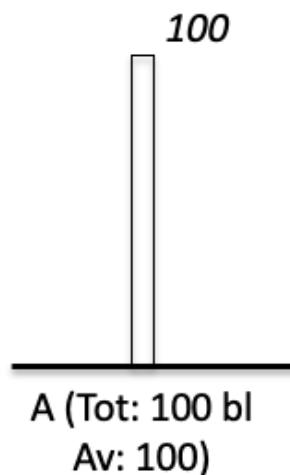
On one view about what makes our lives go best [...] the Century of Ecstasy would have great value for me, this value would be finite, or have an upper limit. In contrast, since each day in the Drab Eternity would have the same small value for me, there would be no limit to the total value for me of this second life. This value must, in the end, be greater than the limited value of the Century of Ecstasy. (Parfit, 2004, pp. 17–18)

In this Chapter, I will here present a variant of one of the oldest impossibility theorems in population axiology, known as Benign Addition Paradox. This is a variant of the Mere Addition Paradox, originally formulated in (Parfit 1986, 419-442) and subject to several variations, such as the ones in (Rachels 2004). The variant called “Benign Addition Paradox”, or sometimes “Up and Down argument” (Parfit 2016, pp. 122), is originally from (Huemer, 2008, pp. 3–5). Since he is the first to propose it in print, I will adopt Huemer’s terminology when referring to this paradox.

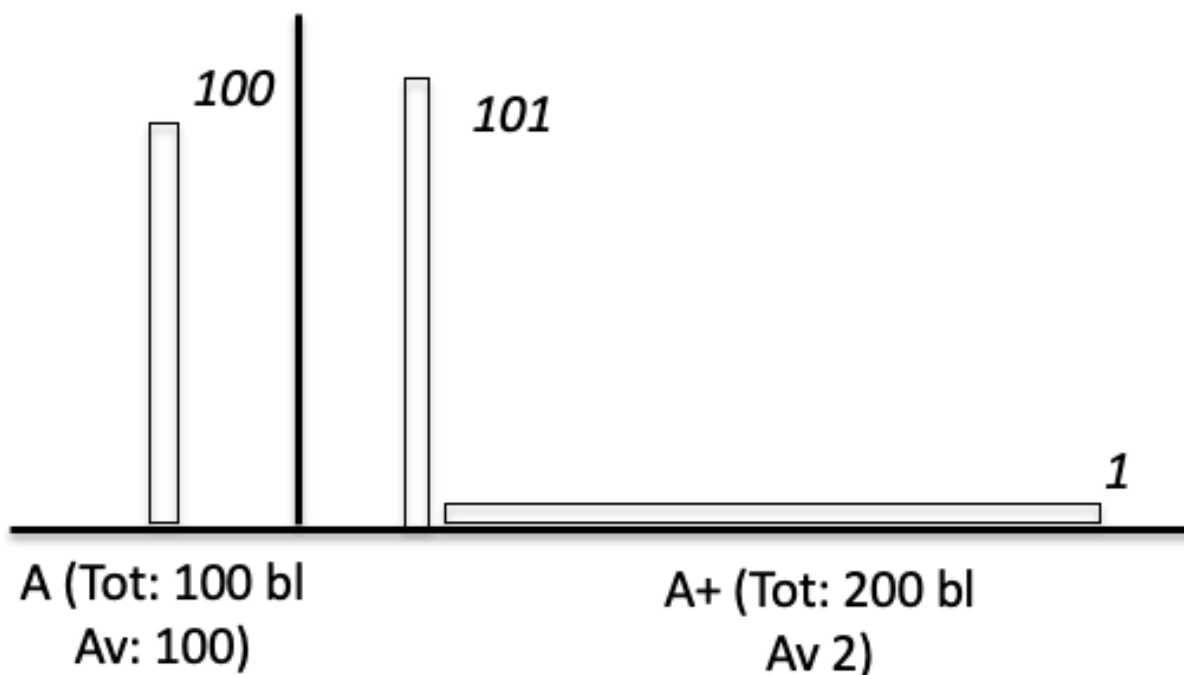
The Benign Addition Paradox is easier to resist than other impossibility theorems (see footnote 15). However, it is representative of what an impossibility theorem in population ethics is, as it highlights all the most relevant tensions that emerge among aggregative theories. I will first explain how the theorem works step by step. Then, I will examine some of the most popular ways people have tried to reject each step, and some of the strongest reasons why we should not be satisfied by any of these attempts. In appendix 2 the reader can find a list of adequacy conditions that we want our theory of population axiology to respect.

2. The Benign Addition Paradox

Consider what we'll call Population A, represented in the graph below. The graph represents a population with 1 billion people. As before, the width of each square is population size, whose precise number of individuals is specified below each population. The height of each population represents wellbeing: I illustrate it with a precise number, visible above the population. The number is purely representational, it does not need to correspond to anything more specific than "wellbeing level". I assume that wellbeing from 100 and above is very high. Below the graph, I specify total and average wellbeing of the population. It will become important later in the argument.



A is a population of 1 billion people, with an average wellbeing of 100, and a total wellbeing of 100 billion. Suppose we could change it as follows: everyone's life improves, and we add to this population 99 billion people, all with a good life. The life of these additional people, while being good, is not as good as the life of the others. Indeed, it is just barely good. Assume that every wellbeing below 5 is very low. Let us say that the life of who was living in A improves by 1, from 100 to 101, and the new lives have wellbeing 1. Call this resulting population A+. I represent it below, close to population A. I will approximate all decimals while averaging.

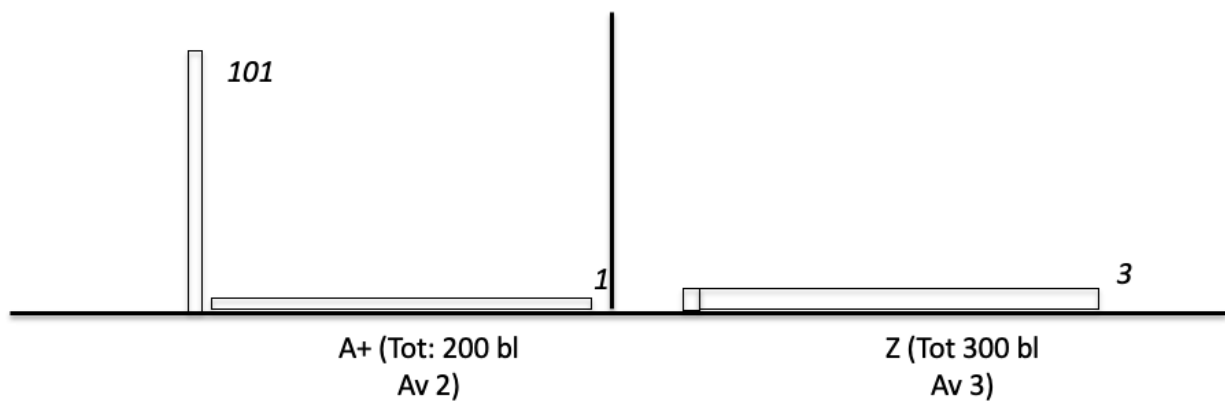


Is the change from A to A+ for the better? On first impression, it may seem that, since the average is higher in A and there is more equality in A, population A is better than population A+. But on reflection, it is quite plausible that A+ is better than A. *Everyone* in population A has it better in A+ rather than in A, and the additional people in A+, who live lives that are good (or at least not bad) would not exist in A. And, A+ has a greater total wellbeing than A. So there is no one for whom A+ is worse than A, and there are some for whom A+ is better than A - perhaps, this is true for everyone in A+.

In a nutshell, people have defended that A+ is better than A due to

Benign Addition. For any two populations P_1 and P_2 , if the wellbeing level of everyone in P_1 is higher in P_2 , and there are additional people with a non-negative wellbeing level, then P_2 is better.

So let us abandon A, and let us focus on the better population A+ (at least according to Benign Addition). Let us now suppose that we can change it again. Suppose we could improve the wellbeing of the people who were added from A to A+, but in order to do so, we need to decrease the wellbeing of people who were in A originally. In doing this operation, however, we would make the population more equal, and we will also be increasing the total and average wellbeing. In particular, we move from population A+ to the population Z illustrated below:



Since population Z has higher average wellbeing, higher total wellbeing, and greater equality than A+, many think that Z is better than A+. It is true that some people's wellbeing is worsened in going from A+ to Z, but many more people's wellbeing is improved in going from A+ to Z.

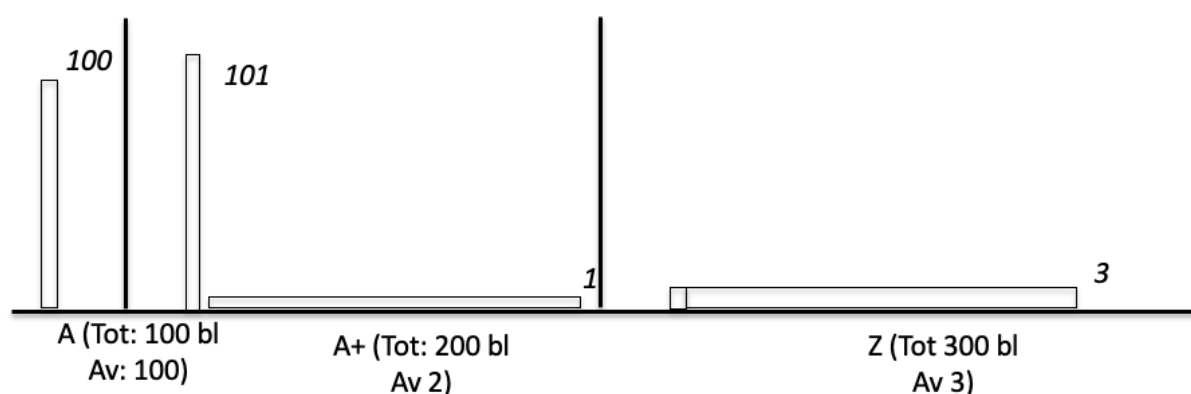
It has been defended that Z is better than A+ due to

Non-Anti-Egalitarianism. For any two populations P_1 and P_2 , if P_2 has higher average wellbeing than P_1 , higher total wellbeing than P_1 , and perfect equality, then P_2 is better than P_1 .

If Non-Anti-Egalitarianism and Benign Addition are true, then Z is better than A+, and A+ is better than A. Many people who accept this conclude that Z is better than A. This is on the basis of

Transitivity of “better than”: If P_1 is better than P_2 and P_2 is better than P_3 , then P_1 is better than P_3 .¹⁷

The graph below shows all populations, so that the reader can see them all at the same time.



However, many people find it incredible that A may be worse than Z. They think that a world with many wonderful lives cannot be worse than a world with lives with very low wellbeing, that are barely good. This implication is known as

The Repugnant Conclusion: For each population A of very many people— say, ten billion— all of whom have a very high quality of life, there is a better population Z consisting of some much larger number of people who would have lives that are barely good. (Adapted from Parfit, 2004, p.10)¹⁸

¹⁷ What this thesis says about the relation “better than” holds for the relation “worse than” too, as I take them to be relations mirroring one another.

¹⁸ Rejecting the Repugnant Conclusion has been argued to be equivalent to accepting

The Quality Condition: There is a perfectly equal population with very high positive welfare which is at least as good as any population with very low positive welfare (Arrhenius, 2000b, p. 41).

So, we have a contradiction. Benign Addition, Non-Anti-Egalitarianism, and Transitivity of “Better Than” imply that Z is better than A. Rejecting the Repugnant Conclusion means accepting that A is better than Z. All four, Benign Addition, Non-Anti-Egalitarianism, Transitivity of Better Than and avoidance of the Repugnant Conclusion, are widely considered very compelling adequacy condition for any theory for aggregating the value of populations. But they are jointly incompatible: no theory can respect all of them. Most of the history of population ethics has focused on developing new attempts to justify why some specific conditions should be rejected rather than another.

Before proceeding, it is worth considering that one may have dismissive reactions after the impossibility theorems. One may think that these theorems prove what some Kantians have been saying all along: that basing our conduct on the consequences of our actions is nonsensical (Lenman, 2000). Or one may think that impossibility theorems in population ethics show that morality is internally inconsistent, and we should become moral skeptics (Arrhenius, 2000b, pp. 200–201; Cowie, 2020). I consider both reactions far too extreme.

Exploring avenues for avoiding or lessening the impact on moral value of dismissing the impossibility theorems is beyond the remit of this thesis. Hence, I will set this exploration aside. A more sensible reaction is trying to find which adequacy condition is more sensible to reject. I cannot possibly summarize all attempts that have been made from (Parfit 1986) onwards. I will, however, give a sketch of some of the most popular accounts, selecting at least one for each rejected condition.

In what remains of this Chapter, I will examine the problems of rejecting Benign Addition, using the “Average Theory” as a paradigmatic example, and of rejecting Non-Anti-Egalitarianism, using Millian superiorities as a case study. An examination of the rejection of further adequacy conditions will be performed in later Chapters. While later Chapters focus on novel objections to

well-known theories, this Chapter surveys existing objections to the rejection of Benign Addition and Non-Anti-Egalitarianism.

3. Average Principle: the cost of rejecting Benign Addition

3.1 The Average Principle

We will first go through the theories rejecting Benign Addition, the idea that if one increases the wellbeing of each individual of a population and adds people whose life is not bad, the resulting population is better. That is to say, these theories reject that the A+ population is better than the A population. To illustrate the problems of these theories, I choose an illustrative, token theory, that has been the default theory in economics until *Reasons and Persons*: its problems are representative of the problems that theories rejecting Benign Addition have. Namely, I will be examining the

Average Principle. A population is better than another iff it has higher average wellbeing (Parfit, 1986, p. 387).

The Average Principle violates Benign Addition. Consider again population A, with 1 billion people having a wellbeing of 100 each, and population A+, with 200 billion people, and an average wellbeing of 2. Since A has higher average wellbeing than A+, A is better than A+, despite the wellbeing level of anyone in A may be higher in A+, and there are many additional people with a positive wellbeing level. Benign Addition is violated.

This principle was considered trivially true by economists (see Parfit, 1986, p. 387), but since the popularization of population ethics after *Reasons and Persons*, it has been largely abandoned. Recently, some people have defended the principle appealing to the widespread inability of population ethics theories to avoid implausible implications, and by pointing out that we commonly

do not think we make the world better by adding people into the world, but this theory remains held by a minority of people (Grill, 2023; Parfit, 1986, p. 387; Pressman, 2015).

It is, however, very useful to examine why it has been abandoned. It is not only because of its violation of Benign Addition: we may occasionally grant that a justification of the violation of Benign Addition may work – we are dealing with impossibility theorems after all, and some adequacy condition has to give. However, in an attempt to violate Benign Addition in a plausible way, the Average Principle runs into two further problems: it violates what are known as the principle of *Separability of Lives*, and the principle of *Non-Sadism*. Both are devastating problems. What is to be noted is that most theories violating Benign Addition thus far proposed, not only the Average Theory, violates either *Separability of Lives* or *Non-Sadism* too (Arrhenius, Forthcoming). We will see an exception in Chapter 5. Analysing the failure of the Average principle is instructive as it shows that violating Benign Addition pushes towards violating further, apparently unrelated adequacy conditions.

3.2 Separability of Lives

The first problem of the Average Principle is what Parfit calls the “Egyptology objection”. Suppose we are wondering whether or not to have an additional child. Again, to simplify, suppose we know that our child will have wellbeing 100. As we are considering whether to have this child, a new study in Egyptology comes up. This study shows that Egyptians used to have a wellbeing level of 10 billion, unobtainable today. This brings the average welfare of humanity substantially above 100. Thus, due to the Egyptology study, it becomes morally worse for us to have a child. But it seems very odd that studies in Egyptology can determine whether or not it is morally better to have a child, because doing so would lower the average welfare of humanity.

To capture the intuition behind the Egyptology objection, we say that a reason to reject the Average Principle is that it violates

Separability of lives. The goodness of conferring some benefit on one person, or of bringing some people into existence, should not depend on how many other people enjoy that benefit or already exist—e.g., on distant planets. (adapted from (Nebel, 2022, p. 11). See also (Blackorby et al., 2005, sec. 5.1.; Broome, 2004, pp. 117–131; Mulgan, 2001; Parfit, 1986, p. 420; Thomas, 2022b))

Many take Separability to be a crucial desideratum for any plausible theory of population ethics, so the violation of separability is sufficient to motivate the abandonment of the Average Principle. However, since I am using the Average Principle as an illustration of the problems that a theory has when rejecting Benign Addition, in the next Section I will examine further undesirable implications of the Average Principle. In fact, some theories rejecting Benign Addition respect Separability, but run into the problems I am about to introduce.

3.3 Non-Sadism

Suppose that, starting from A, we could add very many people (say, 90 billion), whose wellbeing is slightly worse than A (say, wellbeing level 98). Or, we could add a single person who has a bad life (say, with wellbeing -100). The average principle will tell us that adding the bad life is better than adding the many lives slightly worse than lives in A (the average of having 100 billion people with 10 billion at wellbeing 100 and 90 billion at wellbeing 98, is 98.5. The average of having 10 billion people + 1, 10 billion at wellbeing 100, 1 at wellbeing -100, is 99.99). But it cannot be that adding people with a bad life is better than adding people with a good life.¹⁹

We say that Average Principle violates:

¹⁹ Some Average theorists are totalist for negative lives, such as for example (Grill, 2023). They still imply that it is better to add lower level lives than higher level lives.

Non-Sadism. It cannot be better to add bad lives than good lives to a population. (Adapted from (Arrhenius, 2000b, p. 64))

If one considers non-human animals, the fact that it can be bad to create good lives, which is one of the basis for violating Non-Sadism, becomes an explosive problem. In fact, some barely conscious beings will plausibly have, at best, lives that are just barely better than neutral (we will discuss this point more in Chapter 6: Barely good lives). For now, it is enough to note that, as their capability of experiencing wellbeing is limited, barely conscious lives cannot reach much above the neutral level. It has been suggested that non-human species (Arrhenius, 2000a, p. 250; Blackorby et al., 2005, Chapter 11; Tännsjö, 2002, pp. 349–354), such as insects (Sebo, 2023; Williamson, 2021) or oysters (McTaggart & Broad, 1968), may live the just described barely conscious lives.

If the Average Principle is correct, then the birth of many such simple creatures (say, with a new hive) is often intrinsically a tragedy (setting aside the positive consequences that simple creatures may have in an environment).²⁰ Consider a house where there are three awesome humans lives, and an insect whose life is at best barely good is about to procreate. Before the insect gives birth to hundreds of baby insects, the average wellbeing of the house is three quarters of an awesome human life. After the insect gives birth, even if the baby insects are as happy as they could possibly be, the average is a life that is just slightly above barely good: a dramatic drop from being three quarters of an awesome human life, and a genuine tragedy. But we think that the birth of a new hive with happy baby insects is something neutral at worst: this is strongly counterintuitive.

One may think that, to avoid this result, it is enough to postulate that the average must be species-adjusted. That is, population value should be the average percentage of wellbeing that a species can achieve. In the case of the insect giving birth, since all insect offspring were having the best insect life possible, each life should count as an awesome life in insect terms. Thus, before the

²⁰ This point has been firstly made by (Williamson, 2021) against Critical Level Theories, another theory rejecting Benign Addition: to my knowledge, no one has raised this point against the Average Theory.

insect gives birth, the house has four beings with awesome life, and after the insect gives birth, the house has hundreds of beings with awesome lives: the average remains the same.

But species-adjusted averagism has even more unacceptable results. It implies that, for any population of humans where everyone is perfectly happy except for one person, who is *almost* perfectly happy, there is a better population consisting of a *single* barely conscious insect that is perfectly happy in insect-terms. This is even less acceptable than the Repugnant Conclusion.

This is just representative of the struggles that *any* attempt at species-adjusted (and time-adjusted) population axiology has to face (Williamson 2021). The arguments extend to time-adjusted axiologies, according to which the bar where a life is good changes in different eras, so that a barely good life in the Middle Ages would be considered a terrible life today.

The violations of Separability of Lives and of Non-Sadism are to be considered fatal for the Average Principle.²¹ And, most other theories rejecting Benign Addition violate at least one between Separability of Lives and Non-Sadism, too (we will see an exception in Chapter 5). We should abandon these theories, and accept Benign Addition.

Most notably, the violation of Non-Sadism is a problem for so-called Critical Level Theories (Arrhenius, 2000b, pp. 72–87, Forthcoming, pp. 105–128). According to Critical Level Theories, the value of a population is the sum of the wellbeing levels above a certain Critical Level minus the sum of wellbeing levels below the Critical Level (Blackorby et al., 1997, 1998; Bossert et al., 2022; Thornley, 2022). The Critical Level is a higher level than the neutrality level. While I do not examine Critical Level theories in this thesis, I examine two similar theories in the continuation of this thesis. One closely related family of theories, called “critical range theories”, is discussed in Chapter 5: Totalism: Between Repugnance and Greediness. In the next Section, I will examine another closely related family of theories, known as “superiority theories”, that are representative of theories violating Non-Anti-Egalitarianism.

²¹ I am not mentioning Parfit’s “hell three”(Parfit, 1986, p. 422) because it is a violation of Non-Sadism”.

4. 'Superiority' Theories Principle and Non-Anti-Egalitarianism

4.1 Value Superiority

One of the most influential contributions of John Stuart Mill to wellbeing theories is the idea that there may be “superior” and “inferior” kinds of good. By answering an objection to Hedonism present in Plato’s *Philebus*, according to which living the life of Socrates, with all the great accomplishments it contains is certainly better than living the life of a pig that is constantly experiencing pleasure, (Plato, 1975, p. 21 b-d), Mill claims that “some kinds of pleasure [such as Socrates’] are more desirable and more valuable than others [such as the pig’s]”, such that no amount of the less valuable pleasures is better than any amount of the more valuable pleasures (Mill, 2000, p. 11). The more desirable pleasures are part of what we will call “superior” goods, and the less desirable pleasures are part of the “inferior” goods. The idea is that there is a class of “superior” goods and a class of “inferior” good such that any good in the “superior” class is better than any in the “inferior” class.

The idea that the presence of higher and lower goods, not necessarily of hedonist kind, can avoid the Repugnant Conclusion and solve the impossibility theorems of population ethics is widely popular: it has been suggested, among others, in (Carlson, 2007, 2010, 2022; Crisp, 1988, 1992; Edwards, 1979; Glover, 1990; Griffin, 2002; Lemos, 1993; Nebel, 2022; Parfit, 2016; Portmore, 1999; Rachels, 2001; Riley, 1993, 1999, 2008, 2009; Skorupski, 1999; Thomas, 2018). This is not surprising, since Mill elaborates this proposal in the context of solving what we may consider the intrapersonal version of the Repugnant Conclusion.

The standard formulation of views for ranking populations according to “superior” and “inferior” goods is the following. Suppose any life L_n can have some amount of Superior good S_n and some amount of Inferior good I_n . According to

(Strong) Superiority View. A life L_1 is better than another life L_2 iff either

1. $S_1 > S_2$

Or

2. $S_1 = S_2, I_1 > I_2$ (Carlson 2022, Thomas 2018).

This means that, when ranking two populations, a population is better than another if it has more Superior goods: if two populations have the same Superior goods, then we should use the Inferior goods as a tie-breaker. This structure can be formulated in such a way that it includes more than two kinds of goods, and to include “Inferior” and “Superior” bads.

This theory violates Non-Anti-Egalitarianism. Consider again population A+ and Z in the Benign Addition Paradox. A+ has one billion people with welfare 101, plus 99 billion people at welfare 1. Z has 100 billion people at welfare 3. According to Non-Anti-Egalitarianism, Z is better because it has higher total and average welfare, as well as perfect equality.

Superiority views would describe the situation differently. For people to have awesome lives, as people in A+ have, they need to have at least one Superior good. For people to have barely good lives, like lives in Z, they may have no such good. Since A+ has more Superior goods than Z, it is better: Non-Anti-Egalitarianism is violated.

Superior views, however, seem to have an excellent explanation as to why Non-Anti-Egalitarianism is to be violated. These theories agree that, in a sense, the greatest *amount* of good is present in Z rather than in A+. But that is not the amount that matters. Some goods in A+ matter much more than any good in Z. This is why it is not a problem to violate Non-Anti-Egalitarianism.

However, things are not that simple. This justification for violating Non-Anti-Egalitarianism stands and falls with the plausibility of the existence of “inferior” and “superior” goods. And, as it turns out, the existence of these goods is really hard to justify.

The main difficulty is that all good and bad things seem to come in a fairly continuous spectrum, where the goods differ only in degree (as noted in Chapter 2. Assumptions and Methods of Population Axiology). The presence of a continuous spectrum means that, if one good P is better than

another good Q, one can transition from good P to good Q in a series of incremental steps, $X_1, X_2, X_3 \dots X_n$, where each good X_1 to X_n is better than Q but worse than P.

There are various examples of this in axiology. There is a range of pain intensity that goes from the triviality of hangnail pain to the unbearability of torture (see eg. Temkin 2012; see also Section 5 of this Chapter 3). There is a range of artistic quality that goes from excellence of Mozart, to the mediocrity of muzak, maybe going through Haydn and Enya (Parfit, 2004, pp. 26–27). We will see similar examples with the most intuitively “gappy” goods, such as relationships and achievements, in Chapter 9.

No one thinks that the worse elements of the spectrum, such as muzak or hangnail pain, is a “superior” good (or bad). From there, all the other members of the spectra differ merely by degree, and the degree can be arbitrarily small. This means that, at a certain point of the spectrum, there will be a pair of goods that differ only by an arbitrarily small degree such that one is a Superior good, and one an Inferior good. This is implausible. No one has proposed any substantive argument as to why we should consider it plausible that an arbitrarily small difference in degree can be the difference between an “inferior” and “superior” good.

To see more clearly the problem, I will use some numbers to represent wellbeing (see Chapter 2, Section 3 on how to interpret these numbers). Let us suppose that we can represent art enjoyment with numbers, and that there is a cut-off point at 10 such that any piece of music that causes an enjoyment greater than 10 is Superior to any piece of music that causes an enjoyment not greater than 10. Intuitively, there is very little difference between enjoying a piece at a level of enjoyment of 9.9 and enjoying a piece at a level of enjoyment 10.1. This difference can be barely perceptible. And, intuitively, there is instead a great difference between enjoying a piece of music at that is *almost* superior, at 9.9, and not enjoying any music at all, with 0 enjoyment. This difference is very much perceptible. Now, suppose you have one billion listeners at 9.9. On the *Superiority View*, it is better to turn off the music to everyone except one single individual, if this individual starts to listen to the music that gives 10.1 enjoyment rather than 9.9. This is implausible.

This is implausible in two ways. Either this view gives too much weight to the transition across the cut-off point, or it gives too little weight to what happens below the cut-off point. That is to say, the Superiority View is either too sensitive to what happens around the threshold, which I will call to be *Hypersensitive* to the threshold, or not sufficiently sensitive to what happens around the threshold, which I will call to be *Hyposensitive* to the threshold. While these are helpful concepts for illustrating what is disturbing about Superiority views, their avoidance is not considered an adequacy condition for theories of population axiology.

However, hypersensitivity and hyposensitivity are so unpalatable that, while many are in general attracted by the idea of Inferior and Superior goods, no one has been able to indicate, not even suggestively, a wellbeing theory that provides the require justification for Inferior and Superior goods. Some of the Superior theorists, such as (Carlson, 2007, 2010, 2022; Thomas, 2018), have explicitly said that, *if* a plausible wellbeing theory justifies Inferior and Superior goods, this would provide a way out of the paradoxes of population axiology, without committing that such wellbeing theory can be found.

4.2 Weak Superiority

In an influential footnote, James Griffin suggested that changing the Superiority view into a “Weak Superiority” view can be the key to solve population axiology (Griffin, 2002, pp. 89, footnote 27). On the classic, “Strong” Superiority view, any amount of the “Superior” good outweighs any amount of the “Inferior” good. On the “Weak Superiority” view, there is *some* amount of the Superior good that is better than any amount of the Inferior good.

The most natural reading of the Weak Superiority view is that it is identical to the Strongly Superior view, except that if the Superior good is present in insufficient quantity, it counts as an Inferior good. However, many have pointed out that the Weak Superiority view just collapses in the Strong Superiority view under some very reasonable assumptions (Arrhenius, 2005; Arrhenius &

Rabinowicz, 2005, 2015b; Jensen, 2008). These assumptions are Separability, Transitivity of “better than”, and Completeness/Trichotomy. We have already encountered some of them, but it can be helpful to state them again below.

Separability of lives. The goodness of conferring some benefit on one person, or of bringing some people into existence, should not depend on how many other people enjoy that benefit or already exist—e.g., on distant planets, in the past.

Transitivity of “better than”: If P_1 is better than P_2 and P_2 is better than P_3 , then P_1 is better than P_3 .

Completeness/Trichotomy: For any value-bearers X and Y , X is either better than, worse than, or just as good as Y .

We encountered Separability when we were discussing theories violating Benign Addition, such as the Average Theory. Respecting Separability is considered an adequacy condition for theories of population ethics: it would be a huge cost for Weak Superiority theories to violate it.

In this Chapter, I presented and motivated Separability as an adequacy condition for theories in population ethics in Section 3.2, and I introduced Transitivity of “better than” as an adequacy condition in Section 2. In the next Chapter I will further motivate Transitivity of “better than”.

We have encountered Completeness/Trichotomy in Chapter 2, and I presented no argument to accept it. Indeed, I will reject it in Chapter 7: The Structured Range View. However, Completeness/Trichotomy is a very common assumption: it is important to understand why Weak Superiority collapses into Strong Superiority if we assume Completeness/Trichotomy.

I present the argument in (Arrhenius, 2005; Arrhenius & Rabinowicz, 2005, 2015b; Jensen, 2008; Nebel, 2022) as an argument by contradiction: I show that assuming Transitivity of “better

than”, Completeness/Trichotomy, Separability and that Weak Superiority is different than Strong Superiority leads to a contradiction.

Recall population A and population Z in the Repugnant Conclusion: any Superiority view will want to say that population A is better than population Z, because the goods in A-lives are Superior to the goods in Z-lives. Strong Superiority views say that *any number* of A-lives, including one, is better than any number of Z-lives. Assuming Weak Superiority, there is a number x of A-lives that is better than any number of Z-lives. If Weak Superiority is different from Strong Superiority, then the number x of A-lives that is better than any amount of Z-lives has to be greater than one.

If one single A-life is not better than any amount of Z-lives, as any Weak Superiority account different from a Strong Superiority account implies, then, assuming Completeness, there is a number y of Z-lives that is at least as good as one single A-life.

Suppose we have a population consisting of a single A-life. Suppose we could add to this population either another A-life, or y Z-lives. Given Separability, it should not matter who already exists in the population: it is always at least as good to add y Z-lives than one A-life. Thus, (C1) one A-life plus y Z-lives is at least as good as two A-lives.

Suppose now we have a population of y Z-lives, and we can add either a single A-life or y Z-lives. Again, given Separability, it is better to add y Z-lives. Thus, (C2) $2y$ Z-lives are at least as good as one A-life plus y Z-lives. From Transitivity, (C2), and (C1), we conclude that $2y$ Z-lives are at least as good as two A-lives.

We can re-iterate this process for any x times. Eventually, we reach that, for any x , xy Z-lives are at least as good as x A-lives. Thus, it is not true that there is some number x greater than 1 of A-lives that is better than any number of z-lives. Thus, Weak Superiority is compatible with Completeness, Separability and Transitivity only if any amount of A-lives (including 1) is better than any amount of Z-people: this makes Weak Superiority collapse into Strong Superiority. The proof by contradiction is completed. If Weak Superiority just is Strong Superiority, it should be rejected for the same reasons: it is either Hyposensitive or Hypersensitive. Given this, despite the very great

number of people who have pointed at Superiority Views as a likely way out of the impossibility theorems (Carlson, 2007, 2010, 2022; Crisp, 1988, 1992; Edwards, 1979; Glover, 1990; Griffin, 2002; Lemos, 1993; Nebel, 2022; Parfit, 2016; Portmore, 1999; Rachels, 2001; Riley, 1993, 1999, 2008, 2009; Skorupski, 1999; Thomas, 2018), there is no mainstream Superiority View in population ethics.²²

Is the failure to distinguish Superior and Inferior Goods a sufficient basis to accept Non-Anti-Egalitarianism? Maybe. However, remember that Non-Anti-Egalitarianism implies that it can be better to sacrifice very great amounts of good in order to bring a trivial benefit to many people. This is, in fact, what happens in the transition from A+ to Z. This is *prima facie* hard to believe: what is *prima facie* hard to believe can be an interesting implication of a view, but cannot be considered sufficiently strong to be an adequacy condition in population axiology (Thomas, 2018). I will propose a novel argument to reject Non-Anti-Egalitarianism in Chapter 7, Section 3.3.

For now, we accept Non-Anti-Egalitarianism as an adequacy condition. However, we do so only conditionally to the absence of a convincing distinction between Superior and Inferior Goods. In Chapter 7: The Structured Range View I propose a theory, that I call the Structured Range View, that can be interpreted as an improvement over Superiority theories. The Structured Range View violates Non-Anti-Egalitarianism, and possibly Hypersensitivity, and Hyposensitivity but, I argue, does so in a plausible way. The Structured Range View says that there is a range of welfare levels where a life is neither good nor bad: it is neutral. I argue that no amount of neutral items is better than any amount of good items: while this may read like a truism, I show that, if there is a range of neutral welfare levels, it straightforwardly violates Non-Anti-Egalitarianism, and possibly Hypersensitivity, and Hyposensitivity. Nevertheless, none of these alleged adequacy conditions are more plausible than the fact that no amount of neutral items is better than any amount of good item.

²² See (Nebel, 2022) for an interesting attempt. Another extremely promising attempt is Matthew Clark's theory of "continuous superiority", illustrated in his doctoral thesis.

But Chapter 7 needs to wait. Before illustrating the Structured Range View, we need to examine the other adequacy conditions in the Benign Addition Paradox.

Chapter 4: Transitivity of “better than” and inferences

Abstract. We generally think that, for any options A, B, and C, if A is better than B, and B is better than C, then A is better than C. This is known as Transitivity of “better than”. Despite its intuitive appeal, many have suggested that we ought to reject Transitivity. This proves to be a fruitful move, as it avoids many paradoxes in ethics and decision theory. In this chapter, I provide a new argument for Transitivity: without Transitivity, there is no plausible way to handle cases of partial information. The argument goes as follows. Suppose that a sadist forces you to choose between three options: (A) being mocked for two minutes, (B) witnessing the sadist kill one of your loved ones, or (C) having the sadist doing something worse for you than option (B). We have only partial information in that we do not know what option (C) is; we know only that it is worse than (B). Yet it seems clear that, in cases like this, we should choose (A) over (C) – indeed, it would be irrational to believe otherwise. I argue that only Transitivity can explain why we should choose (A) over (C). I show that this problem generalises to theories rejecting “completeness”, or “trichotomy”, unless these theories restrict the rejection of “completeness”, or “trichotomy”, to a range bounded from above and below. I conclude by showing that “Pessimism”, the view according to which no life is better than nonexistence, has to reject “trichotomy” in a way that cannot be restricted to a range bounded from above.

1. Introduction

My father is taller than me. My brother, however, is even taller than my father. This is enough information to conclude that my brother is taller than me due to the transitivity of the relation “taller than” according to which if A is taller than B, and B is taller than C, then A is taller than C.

While transitivity holds for the relation “taller than”, there is debate as to whether it holds for the relation “better than”. According to Transitivity of “better than”, already introduced in previous Chapters, if option A is better than option B, and option B is better than option C, then option A is better than option C. Some people consider accepting this principle a minimum requirement for rationality (Binmore & Voorhoeve, 2003; Broome, 2004; Huemer, 2008; Parfit, 1982, 1986), for example because, if we have a cycle when ranking our options, with A better than B, B better than C, and C better than A, every option we pick will necessarily be worse than another, and it seems irrational to pick an option when there is a better one available. However, there is a growing interest in theories that violate this principle (Friedman, 2009; Goodrich, 2016; Rachels, 2004; Schoenfield, 2014; Snedegar, 2017; Temkin, 1987, 1996, 2012). In fact, transitivity is a background assumption in many of the paradoxes that have dominated philosophical discussion (Arrhenius, 2000b, Forthcoming; Arrow, 1950; Parfit, 1986; Wilkinson, 2022): the hope is that a coherent non-transitive theory would solve these paradoxes. Whether or not transitivity is a requirement for rationality would have an enormous impact on ethics, economics, decision theory, and other fields (Friedman, 2009; Temkin, 1987, p. 186, 2012, pp. 495–497, 514–521).

The most comprehensive argument against non-transitive views are “money-pump” arguments (Gustafsson, 2022a; Gustafsson & Rabinowicz, 2020; Rabinowicz, 2000). These arguments exploit the circles in the rankings of these theories. There are replies to these arguments, and the debate is still open (Gustafsson, 2022a; Temkin, 2012). My argument is independent from “money pump” arguments.

This chapter aims to offer a novel challenge for non-transitive theories. I focus on an underappreciated feature of transitivity: its inferential role in cases of incomplete information. Sometimes it seems clear which option is best or worst to us, even if the only available description of this option is how it compares to a single other option. Non-transitive theorists struggle to explain why this is the case.

Before examining the inferential problem for non-transitive theories, however, it is important to understand what makes non-transitive theories attractive in the first place. It is important to note what follows in our examination: while most of this thesis is an analysis of what is better and what is worse, this Chapter analyzes primarily what we have more reasons to choose and what we have fewer reasons to choose. This is because I am assuming that we have more reasons to choose what is better, and that what is better is what we have more reason to choose (at least when only considerations of betterness are involved: see the discussion of the “teleology” assumption in Chapter 2. I am aware there may be deontological principles that may not be transitive. But I am hoping to bracket my discussion only to cases where betterness is the only moral consideration). I do not defend this assumption in this thesis, but I find it overwhelmingly plausible. This assumption will be important also in Chapter 7.

2. The case for non-transitivity

Consider the following influential case, adapted from (Temkin, 1996, p. 179, 2012).²³

²³ Some may think that incommensurability or vagueness may help evade this kind of spectrum arguments. This is true only if it is a very acute kind of incommensurability or vagueness. Namely, in Temkin's spectrum, the kind of incommensurability that would allow us to evade his argument requires that for some element of the spectrum, i.e. for some pain P of a certain duration, the next element, which is a slightly less intense pain P' of a twice as long duration would not only be incommensurable with the preceding element but would remain incommensurable however much the duration of P' were increased. See (Handfield & Rabinowicz, 2018).

Hangnail for torture. Suppose you are about to suffer two years of torture. You can press a button to make it just perceptibly less intense, but twice as long. Thus, if you press the button the first time, you will suffer four years of slightly less intense torture, if you press it the second time it will be even less intense but lasting eight years, and so on. Each time you could press the button, it is better for you not to press it: for any unpleasant or “negative” experience, no matter what the intensity and duration of that experience, it would be better to have that experience than one that was only a little less intense but twice as long. However, if you press the button sufficiently many times, say, n times, you will suffer only a mild discomfort, such as hangnail pain, for many years.

According to (Temkin, 1996, 2012), it would be better to suffer mild discomfort for the duration of one’s life than to suffer two years of excruciating torture, no matter the length of one’s life. However, this is incompatible with the transitivity of “better than”.²⁴

Each time one can press the button, it would be better not to press it. Thus, if we need to press the button n times to suffer mild discomfort, it would be better to press it $n-1$ times rather than n times. And it would be better to press the button $n-2$ times than $n-1$ times. And so on, until we reach that never pressing the button is better than pressing the button once. If transitivity of “better than” is true, then never pressing the button and suffering the two years of torture is better than pressing it n times and suffering a lifetime of mild discomfort from hangnail pain.

This is considered by non-transitive theorists a paradigmatic counterexample to the transitivity of “better than”: the most plausible way to allow both that pressing the button each time is worse and

²⁴ The case I presented is intransitive within a choice-set. This is just for the sake of brevity: Temkin’s original is across choice-sets, as he asks to make pair-wise comparisons among options along the spectrum rather than having the full spectrum of options. Temkin’s argument works as an argument for intransitivity both within and across choice-sets. The case I am about to propose, too, works both within choice-sets and across choice-sets.

that mild discomfort for a lifetime is better than two years of torture is rejecting the transitivity of “better than”.

3. Inferential transitivity

Consider the following case.

Sadist Kidnapper. You are at the mercy of a sadist, who wants to make you suffer. To further enjoy your struggles, he wants you to choose how you suffer. He gives you the following two options: he will either (A) mock you for five minutes or (B) have you witness him torturing one of your loved ones. As he offers you these options, he thinks it over for a second and says: “no, forget about (B). I won’t do that. The two options between which you need to decide are (A) I mock you for five minutes, or (C) I do to you something that is worse than (B). I will, however, not tell you what (C) consists of.” You know that the sadist is perfectly reliable.

Many people will have the intuition that, between (A) and (C), it is best for you to pick (A). It seems irrational to be indifferent between (A) and (C), and even crazier to pick (C) over (A).

Transitive theories can straightforwardly explain why. Transitivity implies that, if (A) is better than (B), and (B) is better than (C), then (A) is better than (C). This is the intuitively correct result. Indeed, it seems so intuitively self-evident, it would be irrational not only to deny that (A) is better than (C), but even to be indecisive between the two.

Non-transitive views, instead, struggle to explain why we should think that (A) is better than (C). These two options have no ground of comparison beyond their relation to (B). Indeed, I know of no non-transitive view explaining why we should prefer (C) to (A).²⁵ Thus, if we cannot appeal to

²⁵ Some may think that non-transitivists should just have a principle according to which, if A is better than B, and B is much better than C, then A is better than C. But this is incompatible with their response to Temkin’s spectrum, where most plausibly one makes one’s state of affairs *much worse* each time one presses the button.

transitivity to explain why (A) is better than (C), it becomes really hard to imagine how to explain it at all.

If we look at the explanations of why transitivity of “better than” should fail in the first place, justifying why (A) is better than (C) becomes even harder. For example, a widely shared assumption among non-transitivists is what I’ll call *choice-set dependence*: it is the idea that which option is best entirely depends on the set of alternatives available, that it depends on the choice-set (Harney & Khawaja, 2023; Otsuka, 2018; Ross, 2015; Snedegar, 2017; Temkin, 2012, 2015). This means that excluding some options from the set we are choosing from (the choice-set) can change which option we should pick – even if the excluded options are not the options we would pick.²⁶

However, choice-set dependence is not necessary for non-transitive theories (nor sufficient: in Chapter 7 we will see a transitive choice-set dependent theory), and thus choice-set dependence is not necessary for non-transitivists to be paralysed when picking between the options that the Fickle Kidnapper proposes them. Whichever justification they choose, since there is no information about (C), if our theory is choice-set dependent, we have to be indecisive between (A) and (C). But this is irrational, and a strong reason to reject non-transitive theories.

4. Replies

To ensure the correct result in *Sadist Kidnapper*, non-transitive thinkers may try to modify their theories in two ways. First, they can argue that tie-breaking considerations allow for the inference. I call this the “tie-breaker reply”. Second, they can argue that statistical considerations allow for the inference. I call this the “statistical inference reply”. In this Section, we analyse both replies.

²⁶ Choice-set dependence has been justified by pointing out that choices are “essentially comparative”: options differ in many different factors, and “the relevance and significance of the factors for determining an [option]’s value will vary depending on the alternatives with which it is compared” (Temkin 2012, p.229). (See also (Snedegar, 2017, pp. 102–109). Some have suggested that multidimensionality may be sufficient to make the best option relative to our choice-sets (Muñoz, 2023).

4.1 The “Tie-breaker” Reply

Some have suggested that non-transitivists should use tie-breakers to decide. For example, if two options are incommensurable, but there is one extra reason to go one way rather than another, we can use that option as a tie-breaker (for a proposal similar to the “tie-breaker reply”, see (Hare, 2010; Schoenfield, 2014)).

More precisely, non-transitivists may appeal to the

“Tie-breaker” reply

P1. There is a reason in favour of (A), and one against (C).

P2. If there is a reason in favour of (A), and one against (C), then (A) is better than (C).

C. (A) is better than (C).

In fact, as non-transitivists, we may think that we have no way to decide between (C) and (A) as they are, but we have a reason against (C) and one in favour of (A). Namely, the fact that (B) is better than (C) is a reason against (C), and the fact that (A) is better than (B) is a reason in favour of (A). Either of these facts can be used as a tie-breaker, and either tie-breaker delivers the correct result that (A) is better than (C).

This strategy has two problems. The first is that the most widely shared explanation for non-transitive theories is incompatible with P1. In fact, non-transitive theories are typically *choice-set dependent* (Harney & Khawaja, 2023; Otsuka, 2018; Ross, 2015; Snedegar, 2017; Temkin, 2012): they cannot explain what happens in the choice-set where we have only (A) and (C) by appealing to reasons that arise from a different choice-set, such as the one including only (A) and (B) or the one

including only (B) and (C). Thus, it is unclear how non-transitive theories can appeal to the “tie-breaker argument”.

However, even if non-transitive views had a way to justify P1, for example by not committing to choice-set-dependence, they have yet another problem. If having a reason in favour of (A) and one against (C) makes (A) a better choice than (C), then we ought to be indifferent between (A) and (C) if we simply add an equally strong reason in favour of (C) and one against (A). These reasons would balance the reasons in favour of (A) and the reasons against (C) given by their relation with (B).

To see why, suppose that the kidnapper, when presenting you with (A) and (C), tells you that he had in mind two further options, (P) and (Q), that he then decided not to offer. (P) is better than (A), and (Q) is much worse than (C). He gives no further description of these options.

So now, when evaluating (A), that is five minutes of mocking, versus (C), which is worse than torturing one’s loved ones, we have a reason in favour of (A) (it is better than (B)), one against (A) (it is worse than (P)), one against (C) (it is worse than (B)), and one in favour of (C) (it is better than (Q)). If we reply to my case with a “tiebreaker” strategy, these reasons in favour or against each option balance out so that we should be indifferent or conflicted between the two. However, it seems implausible that the absence or presence of (P) and (Q) should make a difference as to whether option (A) is clearly better than (C): not concluding immediately that (A) is better seems deeply irrational.

We conclude that non-transitivists cannot appeal to the “tiebreaker” strategy to justify that (A) is better than (C) does not work. In the next Section, we analyse a strategy where comparisons with (B) count, but comparisons with options like (P) and (Q) do not count.

4.2 Statistical Inference Reply

Non-transitivists can reply that we can get the correct result in the kidnapper case by appealing to the probability that (A) is better than (C). In fact, it may be that transitivity is very often true, but not true in all cases. While the relation “better than” is not transitive, and we cannot deductively infer that (A)

is better than (C) from how (A) and (C) relate to (B), we are still gaining some information. The relations (A) and (C) have with (B) still provides some indication of how likely it is that (A) is better than (C). In particular, in the kidnapper case, it seems very likely that (C) is better than (A) given that (A) is better than (B) and (B) is better than (C). This reply is a claim about how rankings in different choice-sets relate to probability distributions in other choice-sets.

More precisely, the claim is the following. If (A) is better than (B), and (B) is better than (C), then it does not follow that (B) is better than (C), but there is some sufficiently high probability that (A) is better than (C). If there is some sufficiently high probability that (A) is better than (C), then one ought to pick (A) over (C) on a statistical basis.

I call this the

Statistical inference reply: If (A) is better than (B), and (B) is better than (C), then one ought to pick (A) over (C) on a statistical basis.

There are some intuitively plausible justifications of this claim. One may assume choice-set dependence and hold that, while the ranking of outcomes depends *exclusively* on the choice-set they are in, we can gain some information relying on how outcomes rank in other choice-sets. This information is of statistical nature. For example, one can say that, if (A) is better than (B) and (B) is better than some underdescribed option (C), one ought not to conclude that (A) is better than (C), but one can conclude that (A) *probably is* better than (C), so we should pick (A). Indeed, non-transitivists hold that transitivity usually, but not always, holds (Snedegar, 2017, pp. 111–113).

This reply, however, undermines most counterexamples against the transitivity of the relation “better than”. Remember Temkin’s *hangnail for torture* argument. In that argument, starting from two years of torture, we can repeatedly press a button to slightly decrease the intensity of our pain by doubling its duration. It is always worse to press the button, but it is better to press the button enough times to suffer mild discomfort for the duration of one’s life than to suffer two years of torture.

Rejecting transitivity is supposed to help us preserve both the claim that pressing the button is always worse and the claim that it is better to press the button enough times to suffer a mild discomfort for the duration of one's life than suffering two years of torture. Virtually all non-transitive accounts explicitly aim to reach this kind of result in this case and similar ones (Friedman, 2009; Goodrich, 2016; Muñoz, 2023; Parfit, 2016; Rachels, 2004; Ross, 2015; Temkin, 1987, 1996, 2012).

If a non-transitive theory accepts the *Statistical inference reply*, it fails to justify that it is better to press the button enough times to suffer a mild discomfort for the duration of one's life than suffering two years of torture. Recall that not pushing the button is better than pushing it once, and pushing it once is better than pushing it twice, and so on. According to the *Statistical inference reply*, not pressing the button is probably better than pushing the button any number of times, including the n number that would lead to a lifetime of discomfort. Thus, you should always press the button. Accepting the *Statistical inference reply* makes the rejection of transitivity of "better than" insufficient to avoid the problems that motivated the rejection in the first place.

Finally, the *Statistical inference reply* is contrary to common experience. When evaluating (A) and (C), at no point I find myself thinking about probabilities. Nor I have met anyone who told me otherwise. The answer that (A) is better than (C) is immediate: I see no basis to believe that it follows some probabilistic calculation.

I do not think that any of the replies I gave to the Statistical Inference Reply are conclusive: non-transitivists may have counter-replies. However, I think what I argued in this Section has enough force to be deeply skeptical of the Statistical Inference Reply.

To conclude, I find both the *Tie-Breaker Reply* and the *Statistical inference reply* unconvincing. I am unpersuaded that non-transitive theories can justify why we should pick option (A) rather than option (C) in the Fickle Kidnapper case. Being indecisive between the two is irrational. We should discard any theory that does not respect the transitivity of "better than".

5 The Kidnapper and non-Trichotomy views

Thus far, in this Section, I challenged non-transitivity, according to which if A is better than B, and B is better than C, then A may not be better than C. I argued that, if non-transitivity is true, then one should be indecisive between five minutes of mockery and something worse than seeing one's loved ones being tortured. This is irrational. I argued we ought to accept that the relation "better than" is transitive.

Closely related to the rejection of the transitivity of "better than" are "non-trichotomy views". Non-trichotomy views reject "trichotomy", or "completeness", according to which, for any two options, one is better, worse, or as good as another one (Savage, 1972; Von Neumann & Morgenstern, 2007). Crucially, if one option is as good as another one, then a slight improvement in one of the two options will tilt the scale in favour of that option, making it better than the alternative. While trichotomy is widely shared in standard economics, some philosophers have challenged it in different ways, trying to add another value relation to the way lives can be ranked.

There have been several proposals for a value relation where two alternatives are the one neither better, worse, or as good as the other. Some of these proposed relations, such as incommensurability, incomparability and parity (Chang, 2002, 2016; Parfit, 2016; Rabinowicz, 2008, 2009a, 2022a; Thornley, 2022) are similar in two respects.²⁷ First, if one of these relations holds between two options, then the option is not better, not worse, nor as good as another one. Second, they are non-transitive: if one of these relations holds between A and B, and holds between B and C, it may not hold between A and C. From now on, I will use "incommensurability" as a token non-transitive, non-trichotomous relation of this kind: when I mention it, I am using it interchangeably with incomparability and parity.

Here is an example to illustrate how incommensurability works. Jean-Paul Sartre writes that one of his students, during the occupation of Paris in the second World War, asked him for advice in

²⁷ (Hájek & Rabinowicz, 2022) explore the possibility of a solution to population axiology paradoxes based on degrees of incommensurability.

choosing between two options: either stay by his mother, who would otherwise be alone in an occupied Paris, or to fight for the French liberation (Broome, 2004; Sartre, 1973). It seems that no option is better or worse than the other.

Crucially, the two alternatives are not equally good either. If they were equally good, a small improvement on either side would tilt the balance in favour of that side. This is not the case: if the student were to learn that food is slightly better in the French liberation army than he initially expected, this will not be sufficient to tilt the balance against staying by his mother. We say that these non-transitive, non-trichotomous relations are “insensitive to sweetening” (Hare, 2010): the “sweetening” is a small improvement in one of the two options that would be sufficient to make one option better than the other in a case of equality, but is insufficient in the case of incommensurability. When we have incommensurability among options, such as in the case of Sartre’s student, we should either be indifferent or conflicted between these options.

On all incommensurability views, insensitivity to sweetening is justified by the presence of multiple dimensions of value pulling in different directions, some in favour of one alternative, some in favour of the other. To see how multidimensionality can generate insensitivity to sweetening, consider again Sartre’s student’s choice. There seem to be multiple dimensions at play: one is the dimension of patriotism, the other is the dimension of devotion to family. In Sartre’s student’s choice, a small change does not tilt the balance between staying with the mother and joining the French liberation army because the two options are so different in different dimensions (namely, one is much better on the dimension of patriotism while the other is much better on the dimension of family devotion) that a small change in one dimension does not tilt the balance.

Incommensurability views are (but need not to be) compatible with transitivity of “better than”. However, they are still vulnerable to a version of the Fickle Kidnapper if they accept what I call:

Cyclical incommensurability. Let option B be not better, not worse, and not as good as option A, and let B be much better than option C. This is insufficient to conclude that C is worse than A.

Views committed to Cyclical Incommensurability are vulnerable to the following variation of the *Fickle Kidnapper*.

Fickler Kidnapper. You are, once again, at the mercy of a sadist, who wants to make you suffer. To further enjoy your struggles, he wants you to choose how you suffer. He gives you the following two options: he will either (A) mock you for five minutes or (B) have you witness me torturing one of your loved ones. As he offers you these options, he thinks it over for a second and says: “no, forget about (A) and (B). I won’t do any of that. The two options between which you need to decide are (A') I do to you something that is not better, nor worse, but not exactly as good as mocking you for five minutes, or (C) I do to you something that is much worse than torturing one of your loved ones. I will, however, not tell you what (A') or (C) consist in.”

The choice between (A') and (C) strikes us as only slightly less immediately obvious than the choice between (A) and (C) in the original case. Surely, it would be irrational to pick (C) over (A'), and it would be irrational to be indecisive or conflicted between the two.²⁸ Something much worse than the torture of one of our loved ones is certainly also worse than something that is not better, not worse, but not as good as being mocked for five minutes. Being indecisive in *Fickler Kidnapper* is roughly

²⁸ That we should pick (A') over (C) may be less intuitive than we should pick (A) over (C) in the original case, but this is only because incommensurability is not an immediately intuitive relation, and because the complication of replacing (A) with (A') may make things less obvious. Nevertheless, upon reflection, picking (C) over (A) is straightforwardly irrational.

as irrational as being indecisive in *Fickle Kidnapper*: we should reject any view that cannot guide for action in this circumstance.

At first glance, it seems that incommensurability views cannot avoid being indecisive in *Fickler Kidnapper*: indecision is directly implied by Cyclical Incommensurability. On *Fickler Kidnapper*, we have an option (A') incommensurable, thus not better, worse or as good as, an option (A). (A) is much better than (B), and plausibly much better than (C). So we have that (A') is not worse than (A), which is much better than (C). On Cyclical Incommensurability, if an option A is not better than an option B, and B is much worse than an option C then, absent further information, we cannot conclude that C is better than A. Thus, on Cyclical Incommensurability, we cannot conclude that (A') is better than (C).

Fortunately, there is a way out from indecision in *Fickler Sadist* for incommensurability views. The way out is to reject Cyclical Incommensurability, and to restrict incommensurability to a range. That is to say, one should accept what follows. For some difference in the amount of wellbeing between wellbeing any two levels A and B that we call “A much better than B”, and for some difference in the amount of wellbeing between two wellbeing levels C and D that we call “D much worse than C”

Restriction to a range: if option X is incommensurable with option Y, for any option P much better than Y, P is better than X, and for any option Q much worse than Y, Q is worse than X.

This avoids indecision in *Fickler Sadist*. In fact, for some definition of “much better” or “much worse”, (C) is much worse than (A). Thus, due to Restriction to a range, incommensurability theorists ought to choose (A') over (C). We conclude that any rational incommensurability theory accepts Restriction to a range – that is, it has a range of incommensurability bounded from above and below.

Not all incommensurability theorists can accept Restriction to a Range. To use an illustrious example of an incommensurability theory that cannot work with Restriction to a Range, consider the

philosophical view Sidgwick called “Pessimism” (Sidgwick 1981, p.414-415, particularly note 1), according to which no life can be good. Pessimists cannot deny that painless lives with good things are not worse than nonexistence (or at least they have not put forward any argument to that effect). Nor they can admit that these lives are exactly as good as nonexistence, since some painless lives are (much!) better than others. Thus, Pessimists have to admit that nonexistence is incommensurable with all painless lives.

But some painless lives are much better than others. A Blissful Life, full of all the best things in life and with no pain, is better than a Nice Life with no suffering and a lot of underwhelming goods. Indeed, the Blissful Life can be arbitrarily better than the Nice Life, as the amount of goods differentiating the two lives can be arbitrarily large. For Pessimists to deny that the Blissful Life is better than nonexistence, they have to admit that nonexistence is incommensurable with a set of lives whose wellbeing levels are in an extremely large range. More precisely, nonexistence has to be incommensurable with a set of options whose goodness has no upper bound: on Pessimism, for any definition of “much better”, nonexistence is incommensurable with the Nice Life, and there is no Blissful Life much better than Nice Life that is better than nonexistence. This is a straightforward violation of Restriction to a range. This makes the ranking of lives Pessimists are committed to vulnerable to the *Fickler Kidnapper* objection, and therefore incorrect. Or so I argue in the rest of this Chapter.

6 Pessimism, and its rejection

According to Pessimism, lives are typically bad, and no life can be good. Pessimism has a long tradition, tracking back to Greek tragedies of the 5th century BC,²⁹ and has in David Benatar and

²⁹ See (Sophocles, 401 C.E., vv. 1225–1235). I consider it particularly worthwhile to discuss Pessimism. One of the crucial tasks of philosophy is to highlight hidden assumptions in our pattern of thoughts. The assumptions of Pessimism seem to me very similar to the pattern of thought of long episodes of extreme sadness, where each suffering becomes unbearable, and no pleasure seems to be worthwhile. When I am in this pattern of thought, I am convinced I am rational.

Christoph Fehige its most prominent contemporary proponents. According to Pessimists, the bad things in life—like pain and frustration—make it worse for people to live a life rather than not to live it, and the good things in life—like pleasure and satisfaction—can never compensate the bad things in life (Benatar, 2008, 2013, 2017; Fehige, 1988; Schopenhauer, 2020).³⁰ For an overview on the history of pessimism, see (Van Der Lugt, 2021).

This means that, according to Pessimism, living a life can be, at best, equally as good as not living it. A life is as good as not living it when contains nothing bad at all. In the vast majority of cases, lives will contain some bad things, and so will be worse than nonexistence.³¹ For simplicity, I will call “nonexistence” the “state” of never coming into existence: according to Pessimists, lives are often worse than nonexistence, and can never be better. There is a debate as to whether we can meaningfully talk about a state that can never exist (Arrhenius & Rabinowicz, 2010, 2015a; Bykvist, 2007; Greaves & Cusbert, 2022). I do not want to pick a side in this debate in this thesis. If one does not think it makes sense to speak of nonexistence, one can replace the term “nonexistence” with the locution “not to live any life”.

Pessimists are classically challenged on the fact that, if suffering counts against existing, then the good things in life should count in favour of existing, as it happens for example in (Baum, 2008; Doyal, 2007). I, however, will focus on showing that Pessimists are committed to a strange ordering of lives.

It is a known result that, if one endorses Pessimism, one has to admit that there are some lives consisting of only good things that are not better nor worse than nonexistence, but some of these lives are better than some others (Benatar, 2013; C. Brown, 2011, 2017; Mogensen, 2021; Stroppa, 2021).

I have been told that many people feel that way. It may be important to know whether our sad selves get it right or our cheerful selves get it right.

³⁰ Some Pessimist argue also that a life is, empirically, bad (Benatar 2008, 60-93). Engaging with this *a posteriori* version of Pessimism is beyond the scope of this Chapter.

³¹ Note that Pessimists do not defend that all lives are necessarily *worth ending*, but just that they are *a priori* not better than nonexistence.

Some will consider this result strange enough. I here show that Pessimism implies something even stranger: some lives are *much better* than the lives incommensurable with nonexistence, while being *worse* than nonexistence. This violates Restriction to a Range.

6.1 Known strangeness in the Pessimist ranking

Pessimist discussions often describe the condition of having no suffering as having the same value as nonexistence. However, there is very little discussion by Pessimists as to how to compare badless lives that differ in the goods they contain, the only notable exception being (Benatar, 2013).

For example, let *Blissful Life* refer to a life without any suffering, frustrated preferences, or unrealized projects, and with many intense pleasures, strong preferences satisfied, and wonderful projects realized; let *Nice Life* refer to a life without any suffering, frustrated preferences, or unrealized projects but with many mild pleasures, weak preferences satisfied, and trivial projects realized.

Intuitively, the Blissful Life is better than the Nice Life. It seems hard to deny that the higher the quality of goods in a life, the better this life is. For any fixed number of satisfied preferences, it is better that these preferences are strongly held rather than very weakly held, and for any fixed number of pleasures, it is better that they are intense rather than mild. Thus, some badless lives are better than others.

It is very hard to deny that the Blissful Life is better than the Nice Life. While neither contains any bad thing, the Blissful Life has all the good things that are in the Nice Life, but they are more numerous and present to a greater degree. However if, as Pessimists believe, badless lives are as good as nonexistence, then both Blissful Life and Nice Life are as good as nonexistence. This seems to violate transitivity of “better than”. Assuming Transitivity of indifference, according to which if A is better than B and B is as good as C then A is better than C, if the Blissful Life is better than the Nice Life and the Nice Life is as good as nonexistence, then the Blissful Life is better than nonexistence.

But according to Pessimists, the Blissful Life is as good as nonexistence (for similar arguments, see (C. Brown, 2011, 2017; Mogensen, 2021; Stroppa, 2021)).³²

However, Pessimists can reject that the Blissful Life and the Nice Life necessarily need to be better, worse, or as good as nonexistence. Pessimists can say that Blissful Life and Nice Life stand with respect to nonexistence in another relation, that may be incommensurability, incomparability, rough equality, parity, or they may be imprecisely equally good (Chang 2002, particularly footnote 5). For the sake of synthesis, let us call “incommensurability” any relation holding the following: two items X and Y are incommensurable if X is not better, not worse and not as good as Y, but still in some sense comparable. Pessimists may say that, while Blissful Life is better than Nice Life, Blissful Life and Nice Life are incommensurable with nonexistence. This way, they avoid the implication that Blissful Life are better than nonexistence: in fact, transitivity of “better than” is silent about incommensurability. Incommensurability lets Pessimists evade accusations of inconsistency. However, the strangeness of the Pessimist ranking runs deeper than that, and indeed, it implies that incommensurability extends to a range unbounded from above.

6.2 Pessimist is incompatible with Restriction to a Range

Recall that Blissful Life is a life without any suffering, frustrated preferences, or unrealized projects, and with many intense pleasures, strong preferences satisfied, and wonderful projects realized. I now introduce what I call a *Surplus Life*: it is the Blissful Life with the addition of hangnail pain.

Surplus Life is worse than Blissful Life: it is identical in every aspect, except for the hangnail pain. According to Pessimism, Surplus Life is worse than nonexistence, too: suffering is sufficient to

³² Brown and Campbell target Benatar specifically. My argument, however, targets all Pessimist theories. Also, (Rachels, 1998, p. 104) has somewhat similar arguments against the ‘Asymmetry’ - the view that, while it is worse to add lives worth not living to a population, it is not better to add good lives to a population-, but they do not target Pessimism.

make any life bad, and no amount of good things can compensate for the bad things in life in a way that can make life better than nonexistence.

It is, however, not clear how Pessimists should compare Surplus Life and Nice Life. Nice Life has no pain, while Surplus Life has some trivial pain. But the good things in Surplus Life are much better than the good things in Nice Life. There is a respect for which Nice Life is slightly better, and a respect for which Surplus Life is *much* better: surely, all things considered, Surplus Life is better than Nice Life.

Denying that Surplus Life is better than Nice Life strikes many as absurd: it would mean that it would be irrational to pay *any* price, no matter how small, to gain *any* prize, no matter how big. Indeed, some passages of Pessimists suggest that they would consider Surplus Life better than Nice Life. When Schopenhauer explains why good things do not outweigh bad things, he explains that this is due to their insufficiency. That is, Schopenhauer complains that good things are too brief, rare, and mild to outweigh bad things (Schopenhauer, 2020): this seems to implicitly accept that, if they were *not* so brief, rare, and mild, they *would* outweigh the bad things (Woods, 2014, p. 51).

Similarly, there are passages of Benatar that suggest he would consider Surplus Life better than Nice Life. For example, he writes that

There are [...] intrinsic pleasures in a life. These pleasures sometimes occur in the absence of negative mental states, and are best when they do. Intrinsic pleasures can also coexist with the negative ones (so long as the negative states are not of sufficient intensity to undo the pleasure entirely). Neutral states and relief pleasures obviously can also affect the quality of a life. It is better to have a neutral state than a negative one, and if one has a negative state, relief from it (as soon as possible) is better than no relief. Nevertheless, [...] Neutral states and relief pleasures can be valuable only in so far as they displace negative states. [...] Once alive, it is good to have [pleasures], but they are purchased at the cost of life's misfortune—a cost that is quite considerable. (Benatar 2008, p.73)

I read Benatar as saying what follows: while pleasure (and other good things) never constitute an advantage of lives over nonexistence, pleasure (and other good things) can compensate for the bad things that happen in life *once someone is alive*.³³ I could not find anything in Benatar, Schopenhauer, Fehige, nor in any Pessimist writing I know, any argument that contradicts this interpretation.³⁴

In addition, if Pessimists reject that good things can compensate for bad things at least once someone is alive, then Pessimism makes parenthood much more demanding than it has been recognised. Parents should raise their children with suffering prevention as the unique goal. They should, for example, grow their children in a state of coma, or constantly sedate them, to ensure they never know what suffering is, even if they miss out on all the other good things in life. Of course, on Pessimism, those of us who were *not* raised this way have been wronged by their parents, and can rightfully complain about it.

I conclude that, according to Pessimists, Surplus Life is better than Nice Life. There are strong arguments in favour of this claim, this ranking seems implicit in some influential Pessimist writings, and there is no Pessimist argument against this claim.

Note that, if Surplus Life is better than Nice Life, it can be *much* better. In fact, the difference in good over bad things is enormously in favour of Surplus Life – indeed, it can be arbitrarily so. This leads the Pessimists to a strange ranking of states. Living Surplus Life is much better than living Nice Life. Living Nice Life is incommensurable with nonexistence. But nonexistence is *better than* living Surplus Life. Nonexistence is, at the same time, better than Surplus Life, and not better than something much worse than Surplus Life. This means rejecting:

³³ Other interpretations are available. Benatar stresses that the role of pleasures is *only* to “displace negative states” (Benatar 2008, p.73). What he may mean is that the only value of pleasure is that it reliefs from suffering. However, this seems contradicted by the sentence where he says that “pleasures sometimes occur in the absence of negative mental states, and are best when they do” (Benatar 2008, p.73).

³⁴ Fehige, however, does not assign any positive utility to satisfied preferences, which makes his position less vulnerable to my arguments. I find assigning no positive utilities to anything implausible.

Restriction to a range: if option X is incommensurable with option Y, for any option P much better than Y, P is better than X, and for any option Q much worse than Y, Q is worse than X.

Restriction to a range is not something Pessimists can admit. Some painless lives are much better than others. A Blissful Life, full of all the best things in life and with no pain, is better than a Nice Life with no suffering and a lot of underwhelming goods. Indeed, Blissful Life can be much better than Nice Life (for any definition of “much better”), as the amount of goods differentiating the two lives can be arbitrarily large. For Pessimists to deny that Blissful Life is better than nonexistence, they have to admit that nonexistence is incommensurable with a set of lives whose wellbeing levels are in a very large range. That is, nonexistence has to be incommensurable with a set of options whose goodness has no upper bound: on Pessimism, for any definition of “much better”, nonexistence is incommensurable with Nice Life, and there is no Blissful Life much better than Nice Life that is better than nonexistence. This is a straightforward violation of Restriction to a range. Pessimism is irrational.

In this Chapter, we have seen that rejecting Transitivity of “better than”, or accepting Cyclical Incommensurability, is implausible. In fact, it commits us to being indecisive between something that is worse than seeing a loved one being tortured and being mocked for five minutes (or something that is not worse than being mocked for five minutes). We should accept transitivity, and look for a theory of population axiology that rejects some other condition. Incidentally, this calls for a rejection of Pessimism, the philosophical position according to which no life can be good.

Chapter 5: Totalism: Between Repugnance and Greediness

Abstract. This Chapter explores the defense of “totalist” theories, according to which we should rank populations by summing individual wellbeing levels. I explore Broome’s argument according to which these theories face a dilemma. Either there is a single neutral level of wellbeing, which leads to a particularly implausible version of the Repugnant Conclusion is sometimes called Maximal Repugnance, or there is a range of neutral wellbeing levels, which leads to a problem called “Greediness”, according to which a neutral feature can “swallow” a positive or negative feature. According to some “Neutral range” theorists, the presence of incommensurability makes Greediness an acceptable cost to pay. However, I argue that there are additional costs to Greediness, and that these costs are not worth the benefits of “Greedy” theories. Specifically, I point out that version of Greediness, sometimes called “Maximal Greediness”, is a violation of Benign Addition: this is a significant intellectual cost to pay. Furthermore, “Greedy” theories avoid the Repugnant Conclusion and Maximal Repugnance only in an unsatisfactory manner, that is, by saying that Population A is not worse, but not better either, than population Z. Thus, “totalists” have a dilemma where both horns are equally thorny. If there is only one neutral level, total theories imply Maximal Repugnance, which is hard to believe. If there is a range of neutral levels, total theories are Maximally Greedy, which means that they pay the intellectual cost of Benign Addition without getting the payoff of satisfactorily avoiding the Repugnant Conclusion or Maximal Repugnance.

1. Totalism and Maximal Repugnance

When Parfit introduced the Mere Addition Paradox (the first version of the Benign Addition Paradox introduced in Chapter 3. The Benign Addition Paradox), the number of people who accepted the Repugnant Conclusion was quite small. As the name suggests, the conclusion is *prima facie* unpalatable, and subsequent discussion has focused on how to reject it.

However, from the year 2000 onwards, the impossibility theorems for a satisfactory population axiology grew in number and strength (See Chapter 3. The Benign Addition Paradox). All these theorems had in common the rejection of the Repugnant Conclusion as one adequacy condition. As many people have pointed out (eg. Tännsjö, 2020; Zuber et al., 2021), each of these impossibility theorems can be interpreted as an argument in favour of the Repugnant Conclusion. Indeed, these theorems show that any theory (not only “totalist” theories) that accepts some plausible set of adequacy conditions will imply the Repugnant Conclusion. Today, the number of people who accept the Repugnant Conclusion is quite substantial (Adler & Holtug, 2019; Broome, 2004; Holtug, 2017; Huemer, 2008; Ng, 1989; Ryberg, 1996a, 1996b; Spears & Budolfson, 2021; Tännsjö, 1998, 2002, 2020 and many more).³⁵

I myself have defended the Repugnant Conclusion in previous work (Stroppa, 2019). Like many others who have defended it, I find the following very attractive:

Total View: If other things are equal, the best outcome is the one in which there would be the greatest quantity of whatever makes life worth living (Parfit 1986, p. 387).

³⁵ Indeed, people accepted the Repugnant Conclusion even *before* Parfit. The first mention of the Repugnant Conclusion is by Henry Sidgwick, and after careful consideration, he accepted it (Sidgwick, 1981, p. 415-416). Similarly, McTaggart accepted the intrapersonal version of the Repugnant Conclusion (McTaggart & Broad, 1968).

Another way to illustrate the Total View is the following. Assume that there is a “neutral level” of wellbeing, where adding a person to the population does not make the population better or worse (notice that this is different from how I define it in Chapter 2: see footnote 36): on the Total View, the value of a population is the total sum of wellbeing of people above this level, minus the total sum below this level.

The Total View respects all desiderata of the Benign Addition Paradox (and other impossibility theorems) except the rejection of the Repugnant Conclusion. However, if the upshot of the impossibility theorems is that the Repugnant Conclusion is to be accepted, then the Total View respects all the desiderata of the impossibility theorems. Thus, the impossibility theorems can be seen as an argument to accept the Total View (Huemer, 2000).

The Total View implies the Repugnant Conclusion as follows. Suppose that barely good lives have wellbeing 1. If the value of a population is the total sum of wellbeing above the neutral level minus the total sum below the neutral level, then, for any population with total sum of wellbeing equal to X , there is a population consisting of more than $X+1$ barely good lives that is better than the population with wellbeing X .

There are objections to the Total View that go beyond the Repugnant Conclusion. One example is the counterintuitive implication that it may be better to triple the number of currently existing people rather than double the current people’s wellbeing level. Another objection is known as the Very Repugnant Conclusion, according to which, for any population, there is a better population consisting of people living terrible lives and sufficiently many people with lives that are barely good. See (Arrhenius, Forthcoming p.48-81) on this matter. However, the discussion about these objections has been dwarfed by the discussion on the Repugnant Conclusion, and in fact

³⁶ In Chapter 2, I defined the “neutral level” as a level where life is not prudentially good or bad for a person. This is different from the neutral level as defined here I am assuming here. In the next chapter I will justify the assumption that, if someone lives at a level that is not prudentially good or bad for a person, then it does not increase or decrease population value. However, it is important to note that some total theorist, such as John Broome, define the neutral level as the level where adding a person does not increase population value, whether or not this level is prudentially neutral for a person. Thus, for simplicity, in this chapter I will assume that the neutral level is a level where adding a life does not make a population better or worse, without assuming that it corresponds to the level where an individual’s life is neither good nor bad.

assimilated by it: the arguments for accepting or rejecting the Repugnant Conclusion are similar to the arguments for accepting or rejecting these two further objections (Spears & Budolfson, 2021; Tännsjö, 1998).

However, biting the bullet on the Repugnant Conclusion exposes the Total View to another objection. As pointed out by many authors (Arrhenius, Forthcoming; Blackorby et al., 1998; Broome, 2004; Carlson, 1998; Mulgan, 2002; Rachels, 2004 and others) mirror version of the arguments in favour of the Repugnant Conclusion imply

The Reverse Repugnant Conclusion: For each population -A of very many people— say, ten billion—all of whom have a terribly low quality of life, there is a worse population -Z consisting of some much larger number of people who would have lives that are barely bad.

One can accept the Reverse Repugnant Conclusion on the same basis one accepts the Repugnant Conclusion: strong impossibility theorems are showing how inescapable it is, and it may be that the correct description of a “barely bad life” does not make the Reverse Repugnant Conclusion all that repugnant.³⁷ So accepting the Reverse Repugnant Conclusion is as justified as accepting the Repugnant Conclusion.

However, someone who believes in the Total View has to accept *both* conclusions, *and* that lives in the Z population of the Repugnant Conclusion can be just slightly better than the -Z population of the Reverse Repugnant Conclusion. This is a great problem for the Total View: the more unappealing Z-lives look, the less plausible it is to accept the Repugnant Conclusion, and the more appealing lives in -Z look, the less plausible it is to accept the Reverse Repugnant Conclusion.

To see why this is problematic, consider what follows. According to the Total View, lives in Z can differ from lives in -Z by as little as a mosquito bite, or even less. Lives in A can be as awesome

³⁷ We will examine candidates for “barely good” and “barely bad” life, as well as candidates for the “neutral level”, in Chapter 6: Barely good lives

as one likes, and lives in -A can be as awful as one likes. And yet, the difference in value between Z and -Z is more than the difference in value between A and -A. If one considers Z, A, and -A all together, Z is the best population of the set; but if everyone in Z got one single mosquito bite more, Z would be the worst population of the set. This is very hard to believe (Broome, 2004, pp. 213–214, 2009, p. 414; Rabinowicz, 2009a, 2022b; Thornley, 2022).

The conjunction of the Repugnant Conclusion and the Reverse Repugnant Conclusion pose a challenge that Elliot Thornley calls

Maximal Repugnance: Let life x and life y be lives that are identical, except that y has one fewer gumbop's worth of pleasure and one more hangnail's worth of pain than x , such that (1) each population of wonderful lives is worse than some population of x lives and (2) each population of awful lives is better than some population of y lives. (Adapted from (Thornley, 2022, p.11))

This challenge is powerful. While there are people who accept it, I know of no explicit defense of *Maximal Repugnance*.³⁸ Indeed, the few Total View theorists who have discussed *Maximal Repugnance* recognize it as a valid challenge to their view – a bullet to bite only in the absence of better alternatives. However, Total View theorists argue, *there are no better alternatives* (Broome, 2004, pp. 213–214). Total View theorists reluctantly accept Maximal Repugnance, as they believe that all alternatives are even more flawed than Maximal Repugnance is.

Maximal Repugnance is, however, the explicit motivation of some theorists to explore whether we can have a wider zone where wellbeing level are neutral (Gustafsson, 2020; Rabinowicz, 2009a, 2022b; Thornley, 2021). The idea of a “neutral range”, or a “critical range”, will be the topic of the next Section.

³⁸ A defense may be provided by Huemer's intuition debunking argument for the Repugnant Conclusion (Huemer, 2008, p. 907).

2. Critical Range Views

Following the seminal work of (Blackorby et al., 1996, 2005), some have defended that there is not a single level where adding a life is neither good nor bad, but an entire range (Broome, 2004, pp. 175–179; Gustafsson, 2020; Qizilbash, 2007, 2018; Rabinowicz, 2009a, 2022b, 2022c; Thornley, 2022). The presence of this range makes it the case either that some populations are incommensurable with some others, or that it is vague if a population is better or worse than another.

Critical Range Theories are modifications of the Total View that incorporate this range of welfare levels where adding a person does not make the population better or worse.³⁹ On the Total View, there is only one neutral level, and a population is better than another iff it has a greater sum of every individual's welfare above the neutral level minus every individual's welfare below the neutral level. However, Critical Range Theorists point out, there seem to be many equally plausible candidates for where the neutral level is (Rabinowicz, 2009a, p. 406, 2022b, p. 114), or it may be vague which welfare level is neutral (Broome, 2004, pp. 164–185). Thus, we have a range of equally plausible candidates for the zero level. Rankings with different zero levels will disagree as to which population is best: for example, if one population A consists of very many people near the lower bound of the range, and another population B consists of only one person near the upper bound of the range, the ranking where the neutral level is *the* upper bound of the neutral range will consider B better than A, while the ranking where the neutral level is *the* lower bound of the neutral range will consider A better than B.

Thus, on Critical Range Theories, there are many equally plausible zero levels. For this reason, to rank any set of populations, there are many equally plausible preliminary rankings, that differ as to which welfare level is the zero level. On Critical Range Theories, a population is better (worse,

³⁹ Critical Range Views are not to be confused with Critical Level Views, according to which we should count wellbeing positively only above some critical welfare level greater than zero: see (Blackorby et al., 1998, 2005).

equally good) than another if it has greater (lower, equally as much) total wellbeing for each candidate zero level (i.e. each sharpening of the zero level). If a population is better for some rankings but not some other, then the two populations are incommensurable.

Let me introduce some terminology that I will use to describe Critical Level Views. Call a wellbeing level for which it is vague whether it is the neutral level or that is a plausible candidate for the neutral level a “sharpening” of the neutral level. (Broome, 2004, pp. 175–179). Sharpening is a technical term to identify an interpretation, or a precisifications, of a vague term, so it is not originally used for incommensurability, but I will use it for both incommensurability and vagueness in this thesis, for simplicity. If one population has greater total wellbeing than another for any sharpening, this population is better. If one population has lower total wellbeing than another for all sharpenings, it is worse. If one population has the same total wellbeing level as another for all sharpenings, it is equally good. If one population is not better, worse, or as good as another for all sharpenings, the two populations are incommensurate, or it is vague which population is better.

It is important to understand the difference between the “vagueness” and the “incommensurability” interpretation of the Critical Range. To explain it, John Broome writes in his *Reply to Rabinowicz*:

Rabinowicz and I agree that, for some such pairs of worlds, neither world is definitely better than the other, and nor are the two definitely equally good. Let us call these ‘borderline pairs’. Rabinowicz thinks that at least some borderline pairs are incommensurate with each other. On the other hand, I think that no pairs of worlds are incommensurate with each other. I think the relation ‘better than’ is vague, and borderline pairs are instances of vagueness.

What is the difference? When two worlds are incommensurate, each world is definitely not better than the other, and furthermore the two are definitely not equally good. Nothing is indefinite there. But I think there are no cases of incommensurateness. Of any borderline pair of worlds, I think that neither is definitely better than the other, and that neither is definitely

not better than the other, and that the two are not definitely equally as good as each other, and that nor are they definitely not equally as good as each other. (Broome, 2009, p. 415)⁴⁰

There is another distinction between different critical range theories, that is important to recognize even if it will play little role in this thesis. According to some theories, there is only one (contributively) neutral level, but many wellbeing levels are incommensurable or on a par with it (Qizilbash, 2007, 2018; Thornley, 2022), or for many levels it is indeterminate whether they are as good as the one neutral level (Broome, 2004, pp. 175–179). According to others, there are multiple neutral levels that are all neutral (Gustafsson, 2020; Rabinowicz, 2009, 2022b).

Despite being very similar to the Total View, Critical Range Views that deny the existence of a unique neutral level do avoid Maximal Repugnance (whereas the Critical Range Views where there is one zero level, but it is vague, still imply it). According to Maximal Repugnance, there are two arbitrarily large populations Z and $-Z$ that are identical if not for an arbitrarily small amount of pleasure and pain such that a population A of ten million wonderful lives is worse than Z , and a population of $-A$ ten million awful lives is worse than $-Z$. For this to happen, the wellbeing of people

⁴⁰The disagreement between Broome and Rabinowicz as to whether the neutral range is a matter of incommensurability or vagueness is linked to a disagreement about what is known as the “Collapsing Principle” (Broome, 2004, p. 174; Rabinowicz, 2009a, p. 394). On the Collapsing Principle, if y is definitely not Fer than x , but it is not the case that x is definitely not Fer than y , then x is definitely Fer than y (Broome, 1997). This means that, if a population Y is definitely not better than a population X , but Y is definitely worse than X , then Y is worse than X . Broome embraces the Collapsing Principle. If the Collapsing Principle is true, then any case of (definite) incommensurateness is actually a case of vagueness: when we think that two populations are incommensurable, we actually mean that our term “better than” is not sufficiently precise to indicate which population is better or worse, or if they are equally good. There is a long discussion on the Collapsing Principle, with Broome’s view supported for example by (Andersson, 2015; Andersson & Herlitz, 2018; Constantinescu, 2012, 2016), and with arguments against it by authors such as (Carlson, 2013; Elson, 2017; Gustafsson, 2018; Rabinowicz, 2009b) arguing against it. On the basis of the arguments against the Collapsing Principle, Rabinowicz allows for the possibility of incommensurability as distinguished from vagueness I do not intend to take a stand on this important metaphysical issue, nor do I need to. In fact, the theory I propose in Chapter 7, is meant to be compatible with both Broome’s and Rabinowicz’s position.

in Z needs to be positive, and the wellbeing of people $-Z$ needs to be negative. This is not possible on Critical Range Views.

In fact, on Critical Range Views with multiple neutral levels, if two populations Z and $-Z$ have people whose wellbeing differs by very little, there are three possibilities when comparing A , $-A$, Z and $-Z$. First, Z and $-Z$ have positive wellbeing levels for all sharpenings, so $-Z$ is not worse than a population of ten million awful lives. Second, Z and $-Z$ have negative wellbeing levels for all sharpenings, so Z is not better than a population of ten million awesome lives. Third, at least one population between Z and $-Z$ will have all lives in the critical range, so it will be incommensurable with both A and $-A$.

Thus, Critical Range Views admitting more than a neutral level do not imply that there are two arbitrarily large populations Z and $-Z$ that are identical if not for an arbitrarily small amount of pleasure and pain such that a population A of ten million wonderful lives is worse than Z , and a population of $-A$ ten million awful lives is worse than $-Z$. That is, Critical Range Views do not imply Maximal Repugnance.

At least, not strictly speaking. There are two ways we may be unsatisfied by how Critical Range Views do not imply Maximal Repugnance. One will be the focus of Section 4 of this Chapter. The second is that Critical Range View may avoid Maximal Repugnance in a trivial way if the upper and the lower bound of the Critical Range are too close together. In fact, Critical Range Views still imply:

Weak Repugnant Conclusion: For any population, there is a better population consisting of people just above the critical range.

And:

Weak Reverse Repugnant Conclusion: For any population, there is a worse population consisting of people just below the critical range.

Critical Range Views are plausible only if the upper bound of the Critical Range is sufficiently low to make the Weak Repugnant Conclusion unproblematic, and the lower bound of the Critical Range is sufficiently low to make the Weak Reverse Repugnant Conclusion unproblematic.

There is no exploration of what would make a Weak Reverse Repugnant Conclusion unproblematic. I will not attempt this exploration in this thesis. Instead, in Chapter 6: Barely good lives I will give an overview of what population ethicists consider an acceptable form of the Weak Repugnant Conclusion. This will provide guidance to any Critical Range View that wants an acceptable Weak Repugnant and Reverse Repugnant Conclusion.

Before exploring what size should the range have, however, it is important to introduce the main problem with Critical Range Views. In the next Section, I introduce the concept of “Greediness”, that Broome identifies as a major challenge to Critical Range Views. Since some Critical Range theorists, most notably Rabinowicz, have important replies to the Greediness worry, I will illustrate further problems with Critical Range Views in Section 4.

3. Greediness in Broome

Broome (2004, pg. 148/170) illustrates the Greediness objection to Critical Range Theories with a version of the Mere Addition Paradox (an earlier version of the Benign Addition Paradox). Suppose the critical range is from 0 to 10 (i.e. 0-10 are plausible candidates for the “zero level”). Broome considers four populations, with wellbeing distributions as follows:

Population A: 4, 4, 4, ... 4, 6, does not exist

Population B: 4, 4, 4, ... 4, 6, 1

Population C: 4, 4, 4, ... 4, 4, 4

Population D: 4, 4, 4, ... 4, 4, does not exist

According to Critical Range theories, A is incommensurable with B, C is better than B, D incommensurable with (or vaguely as good as) C, and D is worse than A. None of this is contradictory, per se. However, Broome points out that

Moving from *A* to *C* involves two things. First, the second-last person's wellbeing is reduced from 6 to 4. This is a bad thing. Second, an extra person is added at level 4. This is a neutral thing. The net effect of one bad thing and one neutral thing should be bad. But according to our theory, it is not bad; it is neutral. Incommensurateness is not neutrality as it intuitively should be. It is a sort of greedy neutrality, which is capable of swallowing up badness or goodness and neutralizing it. (2004, 170).

On several separate occasions Broome shows that this “greedy neutrality” has important, unacceptable practical implications when evaluating what to do with respect to climate change (Broome, 2004, 2005, pp. 407–411, 2022, pp. 149–150). The example I report below is from the most recent of these papers.

Broome asks us to “let the goodness of the best neutral life be 40 and the goodness of the worst neutral life 10” (2022, pg. 149-150). Consider the following two populations, that may be the effect of different policies for climate change.

No climate change: 10 billion people at welfare level 23

Climate change: 8 billion people at welfare level 20

Recall that, on Critical Range Theories, if different sharpenings of the neutral level disagree as to which of two populations is better, then the two populations are incommensurable. Since we assume that the critical range is from 40 to 10, the two populations are incommensurable.

In fact, on the “sharpening” where the neutral level is 40, the “climate change” population is best, as its value is 8 billion times -20, while the value of the “No climate change” population is 10 billion times -17, and $-20 \times 8bl > -17 \times 10bl$.

On the contrary, on the “sharpening” where the neutral level is 10, the “No climate change” population is best, as its value is 10 billion times +13, while the value of the “Climate change population” is 8 billion times +10, and $13 \times 10bl > 10 \times 8bl$.

So, on Critical Range Theories, the two populations are incommensurate.

But this is plainly the wrong conclusion. The distribution with climate change is worse because it is worse for every person who lives. That is a bad thing. It also has a smaller population, which is supposed to be a neutral thing. A bad thing together with a neutral thing should be a bad thing. But the way the theory works, the supposedly neutral change swallows up the bad change and neutralizes the two changes together. (Broome 2022, pg. 150).

We can capture Broome’s “Greediness argument” with the following adequacy condition:

Non-Greediness: “The net effect of one bad thing and one neutral thing should be bad” (2004 p. 170)

However, Critical Range Theorists such as Rabinowicz complain that Broome’s argument seems to rest on an interpretation of neutrality that they do not share. In particular, Rabinowicz (2009a) points

out what follows. Broome seems to interpret neutrality as something that, on its own, makes no changes to population value. However, another plausible interpretation of neutrality, which Broome himself appeals to at points (eg. 2004, pg. 176), is that something neutral does not make things better or worse. This second interpretation of neutrality “does not imply that [neutral] changes don’t “count against other values” and that “they can simply be ignored in the total evaluation of outcomes”: on this interpretation, neutral elements are to be taken into account when evaluating an outcome, and can make the comparison between the outcomes vaguer, or can make outcomes incommensurable (Rabinowicz 2009a, p. 399). And indeed, this is the kind of neutrality that Critical Range Theorists are trying to capture.

If the kind of neutrality Critical Range Theorists are trying to capture can count against other values, “There needn’t be anything wrong with greedy neutrality” (Rabinowicz 2009a, p. 399). When it comes to large scale events that will have some bad consequences for some individuals but also impact population size, such as Broome’s aforementioned example involving climate change, there is nothing strange in believing that “total effect of such large-scale events might well turn out to be neither good nor bad, but incommensurate” (Rabinowicz 2009a, p. 399). Gustafsson presents a similar position, accepting that Greediness is just part of the package of any view based on incommensurability (Gustafsson, 2020, 2023).

While I think this reply has some force, I think more needs to be said before accepting Greediness. This is because there is an instance of Greediness, that Elliot Thornley calls Maximal Greediness, that I show is a violation of Benign Addition. Additionally, the payoff of this violation is only an unsatisfying avoidance of the Repugnant Conclusion (and consequently of Maximal Repugnance), according to which A is not worse than Z, but not better either. I do not think that this unsatisfying avoidance of the Repugnant Conclusion is worth the intellectual cost of giving up Benign Addition.

4. Maximal Greediness and the Benign Addition Paradox

Consider the Benign Addition Paradox again. Consider population A of the Benign Addition Paradox, with a billion people living wonderful lives, at what we called wellbeing level 100. Let us suppose that the Critical Range is between wellbeing levels 1 and 10. Suppose we can change our population in the following manner.

We can add 999 billion people at welfare 9, and *double* the wellbeing of people in A. To stress the analogy with the problem of Benign Addition, I call this new population A⁺⁺. Everyone in A fares much better in A⁺⁺. In addition, there are 999 billion people whose lives are very close to the upper bound of the Critical Level: their lives are incommensurable with not coming into existence, and among lives that are incommensurable with not coming into existence, these are among the best lives one can live. Thus the change from A to A⁺⁺ is a change for *much* the better for some people, and a change that is not worse for everyone else. A change for much the better joined with a change that is not worse, intuitively, should make things better. Indeed, it seems hard to deny that A⁺⁺ is better than A. This is what is prescribed by

Benign Addition. For any two populations P_1 and P_2 , if the wellbeing level of everyone in P_1 is higher in P_2 , and there are additional people with a non-negative wellbeing level, then P_2 is better.

This is not what Critical Range Views conclude. On Critical Range Views, A⁺⁺ is incommensurable with A. In fact, while A⁺⁺ is much better than A for most sharpenings, it is not better for all sharpenings. Indeed, in the sharpening where we are considering the wellbeing level 10 as our 0 level, A is better than A⁺⁺: on this sharpening, the wellbeing 9 of the 999 billion people produces a negative of 999 billion, which outweighs the total of 190 billion of positive wellbeing that is enjoyed by the single billion people who is living both in A and in A⁺⁺. Since the sharpenings

disagree as to which population is better, the two populations are incommensurable, even everyone who lives in A fares better in A++, and in A++ there are additional people with non-negative lives. Critical Range Views are in contradiction with Benign Addition.

The reason why Critical Range theorists reject that A++ is better than A is that they reject that a change for much the better joined with a change that is not worse makes things better. In Section 5 of Chapter 4: Transitivity of “better than” and inferences, where we discussed sensitivity to “sweetening”, we have seen that incommensurability may justify that a change for the better and one for the not worse necessarily result in something better. However, the incommensurability of Critical Range theories justifies that *any* improvement, no matter how big, joined with a change that is not worse makes things not worse: this is a quite extreme case of “sweetening”, that I expect many to find unpalatable.

To see why the change from A to A++ can be a change for the much better joined with a change that is not worse, consider what follows. In the example I used, the change from A to A++ involves doubling the welfare of people in A. However, the welfare of people in A can be increased by any arbitrarily large amount, and so long as enough people within the range are added to A, the resulting population A++ will be incommensurable with A.

If one considers negative lives, we get a problem similar to the just highlighted rejection of Benign Addition. A mirror argument from the one just presented concludes that, for Critical Range Views, a population -A where one billion people are living awful lives at welfare -100 is incommensurable with a population ---A where one billion people are living much more awful lives at welfare -200 (or any arbitrarily worse welfare level) but there are 999 billion people (or arbitrarily more) living at welfare +1. The billion people living in both populations fares much worse in ---A than in -A, and the additional 999 billion people live lives that are not better, not worse, nor as good as nonexistence. A change for the much worse and a change that is not better, not worse, nor equally good should combine for a change for the worse overall. But, on Critical Range Theories, -A is

incommensurable with ---A, since the sharpening where we consider 0 the neutral level gives a greater total value to ---A than to -A.

As we have seen in the last Section, Broome considers similar cases, albeit all within the “critical range”. What Broome finds problematic about this case is that a change for the worse and a neutral change result in a neutral change: he calls it a problem of Greediness. Thornley finds that we have a particular case of Greediness when A^{++} is incommensurable with A, and ---A is incommensurable with -A. Thornley calls it

Maximal Greediness: for any population of awful lives and any population of wonderful lives, (1) there is some population of straightforwardly-better-than-blank lives such that the population of awful lives is not worse than the population of wonderful lives plus the straightforwardly- better-than-blank lives, or (2) there is some population of straightforwardly- worse-than-blank lives such that the population of wonderful lives is not better than the population of awful lives plus the straightforwardly-worse-than-blank lives. (Thorley 2021, pg.17).

This specification of Greediness is a weaker adequacy condition than Broome’s original. It allows that neutrality can “swallow” some goodness and badness, but rules out that it can do so in extreme cases. Nevertheless, the reply by Rabinowicz and Gustafsson works equally well against both Broome’s definition and Thornley’s definition. They can still reply that “there is no reason to expect that adding things that are neutral in this sense will have no neutralizing effects on bad or good things that are being added at the same time.” (Rabinowicz, 2009a, p. 399). However, I have shown that Maximal Greediness is actually a violation of Benign Addition. This is a very compelling adequacy condition: appealing to features of incommensurability may mitigate how intellectually costly a rejection of Benign Addition is but, it seems to me, it remains a non-negligible cost. To accept Critical

Range Views, we need to check whether this intellectual cost is justified by these view's theoretical virtues.

The main virtue that Critical Range Views aim to have over the Total View is the avoidance of Maximal Repugnance, that is a close conjunction of the Repugnant Conclusion and the Reverse Repugnant Conclusion. If Critical Range Views avoid Maximal Repugnance well enough, we may be willing to accept the intellectual cost of Benign Addition. Unfortunately, the way they avoid Maximal Repugnance is not very satisfying: Critical Range Views and The Total View are identical when it comes to the Reverse Repugnant Conclusion, and the way they avoid the "positive" Repugnant Conclusion is not very satisfying. Particularly, as they admit, their theory entails that A is merely *not worse* than Z (Rabinowicz, 2022b, p. 124).⁴¹

To see why, assume again that the critical range is from 0 to 10. Call A a population of a number x of awesome lives, with wellbeing level 100. Call Z a population consisting of $x \times 101$ lives that are just one drop of discomfort away from having a bad life, with wellbeing level 1. A has greater total wellbeing than Z on any sharpening where neutral wellbeing is greater than 1. Conversely, A has lower total wellbeing than Z for a sharpening where neutral wellbeing is 0. Thus, A and Z are incommensurable. A is not better than Z. It is true that Z is not better than A either, but if we want to convincingly avoid the Repugnant Conclusion, we want to say that A is better than Z, not merely that it is incommensurable with it.

Thus, while Critical Range Views avoid Maximal Repugnance, they imply

Weak Maximal Repugnance: Let life x and life y be lives that are identical, except that y has one fewer gumdrop's worth of pleasure and one more hangnail's worth of pain than x , such that (1) each population of wonderful lives is *not better* than some population of x lives and (2) each population of awful lives is better than some population of y lives.

⁴¹ Some Critical Range Theories, such as (Blackorby et al., 1996), have the lower bound greater than 0: I assume they are implausible, as they imply the Sadistic Conclusion and a worse Reverse Repugnant Conclusion.

Critical Range Theorists may, again, reply that this is all part of accepting the incommensurability package. Or they may consider the rejection of Benign Addition a reasonable bullet to bite. Or, they may complain that the best interpretations of all three adequacy conditions should be formulated in terms of “not worse than” rather than “better than”. I am not averse to exploring properties of value relations, even if this leads us to completely rethink how we should structure value.

Nevertheless, if one were to propose a Critical Range View that does *not* imply Maximal Greediness, does not imply Weak Maximal Repugnance, and respects Benign Addition, these would surely all be points in favour of this theory. I will propose such variant in Chapter 7: The Structured Range View.

5. Conclusion: the dilemma between Maximal Greediness and Maximal Repugnance

Maximal Repugnance and Maximal Greediness are two serious problems. We have seen that the motivation behind Critical Range Theories is to avoid Maximal Repugnance. Conversely, Greediness is enough for Broome to abandon Critical Range Theories: he believes that there is one single level that is as good as nonexistence, despite its location being vague. He is aware this commits him to Maximal Repugnance. Broome shows himself to be torn between Greediness and Maximal Repugnance, albeit he seems more inclined to accept the latter. He writes:

I find [Greediness] incredible. For this reason, I think we have to give up the intuition of [a Critical Range]. I think we must recognize that there is only one neutral level: there is only one level of wellbeing such that it is neither good nor bad to add to the population a person whose wellbeing will be at that level. Adding a person at a higher level is good; adding a person at a lower level is bad. True, this conclusion, too, has some counterintuitive consequences [such as Maximal Repugnance]. To mitigate them, I suggested in *Weighing*

Lives that it is vague what the neutral level is. However [...], if this vagueness is extensive, it will be [...] greedy[.] For instance, it will lead to the conclusion that it is not definitely good to control climate change [...]. To mitigate this problem, I suggested that the vagueness of the neutral level must not be extensive. We need a compromise. There must be enough vagueness to mitigate [problems such as Maximal Repugnance], but not enough to immerse us too deeply into the problem of greediness. I do not pretend any compromise will be ideal. (Broome, 2009, p. 414).

Rabinowicz and Gustafsson do not show in print to be as conflicted as Broome when it comes to choosing between Maximal Greediness and Maximal Repugnance. However, I have shown that Maximal Greediness is a significant intellectual cost, as it is a rejection of Benign Addition. I have also expressed some serious doubt that this cost is worth it, as the Repugnant Conclusion is avoided only in an unsatisfying way, and so is Maximal Repugnance.

Of course, Critical Range Theorists may still think that it is better to pay this cost than to fully accept Maximal Repugnance. However, this means to agree with Broome that they are facing a dilemma: on one hand, rejecting Maximal Greediness, as Broome does (followed by many others), and on the other, rejecting Maximal Repugnance, as Rabinowicz does (followed by many others). In Chapter 7: The Structured Range View I will argue that the dilemma between Maximal Greediness and Maximal Repugnance is actually a trilemma: I introduce a new Critical Range View, that avoids both Maximal Greediness and Maximal Repugnance. However, in population ethics, the bedsheet is always too short, and there is always a problem: my theory rejects Non-Anti-Egalitarianism. I will argue it is a cost we should be willing to pay.

Whichever way we want to go, we need to know which lives are just above neutrality. In fact, any “totalists” want that the Repugnant Conclusion they commit to has acceptable candidates for Z-lives, be these Z-lives above a single neutral level or above an entire neutral range. In the next Section, I investigate which lives are considered “barely good”, and which would be the kind of “barely good

lives” that philosophers do not admit can ever be better than A-lives, no matter how many they are, and which lives, instead, would be acceptable to admit as better than A-lives if they are in sufficient quantity.

Chapter 6: Barely good lives and the Intuition of Neutrality

Abstract This Chapter argues that a version of the “Intuition of Neutrality” is supported by the most prominent theories about what a (prudentially) “good life” is. According to the Intuition of Neutrality, there is a “neutral” range of wellbeing levels where it is neither better nor worse to bring a life into existence. I show that, on prominent theories about good lives, this neutral range is sufficiently wide to include some putative examples of “barely good lives” that would make population Z of the Repugnant Conclusion unacceptable: this supports Critical Range Views. My argument proceeds in three steps. First, after examining three influential criteria that have been proposed to identify good lives, which I respectively call “Camus’ criterion”, the “continuation criterion”, and the “better than neutral criterion”, I derive necessary and sufficient conditions for a good life, a bad life, and a neutral life (I call them “Tradeability” conditions). Second, I consider some examples of “barely good lives” in population ethics. I show that some philosophers, including opposers of the Repugnant Conclusion like Parfit, are willing to accept Repugnant Conclusions if the Z-population is composed of some lives, called “Rollercoaster Lives” and “Very Short Lives”, but not others, called “Drab Lives” and “Barely Conscious Lives”. Third, I show that “Drab Lives” and “Barely Conscious Lives” satisfy the conditions for neutral lives, and so do lives slightly better and slightly worse than them. Thus, we have a range of lives whose wellbeing is neutral: it is neither good or bad to bring these lives into existence, and they do not increase population value.

1. Introduction to barely good lives

What is a good life? How does it differ from a bad one? This question is crucial to understanding some important matters in bioethics, choices about how to live one's life, choosing between climate policies, and much more. For example, in order to take some decisions concerning abortion or euthanasia, or understanding whether to have children during a climate crisis, we need to know whether the life of the people involved will be worth living.

As we have seen in the last Chapter, for the purposes of population axiology, understanding what is a good life has a further theoretical purpose. We are trying to understand what it would take for a version of the Total View, or of a Critical Range View, to successfully avoid a problematic version of the Repugnant Conclusion. In order to do that, we need to perform two tasks.

The first task is to understand what are the plausible candidates for necessary and sufficient conditions for a good life. If we identify them, we will be able to recognise which life is definitely above the neutral range. And, if all lives just above the neutral range are the kind of lives that would constitute an acceptable Z-population in the Repugnant Conclusion, then we can accept theories that imply the Repugnant Conclusion.

Which brings us to the next task. Namely, identifying which lives philosophers have indicated as candidate Z-lives that would make the Repugnant Conclusion acceptable, and which lives philosophers have indicated as candidate Z-lives that would not make the Repugnant Conclusion acceptable.

In order to perform the first task, I overview some definitions of what it is to have a life that is (barely) good. This overview will consider different traditions that have tried to identify the necessary and sufficient conditions for a good life. Namely, I will examine Camus' theory according to which a life is good iff it is rational not to commit suicide (subsection 2.2), the Continuation Criterion according to which a life is good iff it is good to continue it at the same momentary wellbeing level (subsection 2.3), and finally the standard population ethics criterion according to which a life is good iff better than neutral (subsection 2.4). I argue that this overview supports that,

for a life to be good, it is necessary sufficient that, by doubling its duration and slightly decreasing its momentary wellbeing, it gets better. I call this the “Tradeability” test for a good life. Similar tests can be performed to identify bad and neutral lives.

Once we have the necessary and sufficient conditions for a good life, I give an overview of lives that population ethicists have considered examples of barely good lives (Section 3). Among these lives, lives consisting of “ups and downs” (called Rollercoaster Lives) and excellent short lives (called Very Short Lives) are considered better than underwhelming lives (Drab Lives) or lives of barely conscious beings (Barely Conscious Lives). I show in Appendix 1 that, while some theories have tried to explain why we consider the former lives better than the latter lives, these theories fail.

Once we have identified what several traditions recommend as sufficient reasons for spotting positive, negative, and neutral lives, we can check whether these putative “barely good lives” are actually good. In Section 5, I point out that not all of them are. Indeed, while Very Short Lives and Rollercoaster Lives are indeed good, Drab Lives and Barely Conscious Lives are neutral. Even more interestingly, lives slightly better and worse than Drab Lives and Barely Conscious Lives are also neutral. Thus, there is a range of neutral lives.

This supports a version of what is known as the Intuition of Neutrality, introduced in Section 4. Many think that there is more than one wellbeing level such that where we do not improve a population by adding a person at this level. This has been called the Intuition of Neutrality (Broome, 2004, 2009; Narveson, 1973; Qizilbash, 2018; Rabinowicz, 2009a).

While some interpret the “neutral range” of Intuition of Neutrality as covering the range of all lives with non-negative welfare levels, this “broad” interpretation is commonly considered too extreme. On a “narrow”, more plausible interpretation of the “neutral range” of the Intuition of Neutrality, above a certain wellbeing threshold adding lives is better, below a certain wellbeing threshold adding people is bad, and between the thresholds adding people does not make things better or worse. This is a version of the Critical Range position illustrated in Chapter 5: Totalism: Between Repugnance and Greediness We saw that classical Critical Range theories with a neutral range run

into the problems of Maximal Greediness, while avoiding Maximal Repugnance in an unsatisfying way. However, the arguments of this Chapter support the presence of such a neutral range: we need to find a Critical Range view without Greediness problems, and that does not violate so many adequacy conditions.

Before starting the examinations of this Chapter, there is an important element to clarify. There are two senses in which a life at a certain wellbeing level can be good (bad, neutral): it can be good because it is good for a person living at this level, or it can be good because it increases population value. I will here be assuming that a life's wellbeing level is good (bad, neutral) for who lives it iff it makes a population better. Not everyone accepts this assumption. Namely, critical level theorists reject this assumption: they think that some neutral lives at the personal level are bad at the general level. But this violates Non-Sadism: see pp. 49/50. To avoid Non-Sadism and related problems, I endorse this assumption.

Other people who reject this assumption are those who accept the "broad" intuition of neutrality (to be introduced in section 4 of this chapter), according to which would say that lives that are good in a personal sense are neutral from the point of view of population value. I reject these theories for reasons I explain in section 4 of this chapter. I know of no other reasons to deny that, if a life is good (bad, neutral) for who lives it, then it makes a population better. Thus, I will endorse this assumption for the rest of the thesis.

2. Barely good lives

2.1 Necessary and sufficient conditions for a barely good life

Philosophers have described lives that are barely good in several ways. Some have tried to provide necessary and sufficient conditions for a barely good life. Some particularly influential attempts of this kind are examined in this Section. Some others have attempted to give examples of barely good lives: the most influential examples are examined in Section 3. As we will see, people have been

more willing to accept a Repugnant Conclusion in cases of complex human lives than in any other case. I will conclude that all the most promising theories share a common intuition: a life is good (bad, neutral) for who lives it if more of it is better (worse, equally good) for who lives it.

I examine three kinds of theories for the necessary and sufficient conditions of a good life. I present them from what I take to be closer to the most immediate interpretation of the term “good life” among non-philosophers, and finish with the more technical definitions. Namely, I examine different variations of *Camus’ criterion*, according to which one’s life is barely good iff it is the worst of the lives where one would not rationally commit suicide; I then examine different variations of the *continuation criterion*, according to which a life is barely good iff it is the worst of the lives that one would continue, and I conclude with the criterion proposed in population ethics, namely that a barely good life is barely better than nonexistence, and that a life is barely good iff it is just better than a life that is as good as a neutral welfare component.

I do not limit my discussion of good lives to population ethics definitions because it is important to see what is unique in the population ethics analysis when compared to other ways of thinking about good and bad lives. Population ethics definitions of a barely good life commonly identify the *neutral level* of wellbeing first (sometimes identified as the level(s) where life is neither better nor worse than nonexistence) and *then* the good life. Instead, most definitions outside the population ethics literature point at the *internal features* of lives. What is interesting is that the internal feature that seems to underlie all these definitions outside population ethics is an idea that will be very familiar to population ethicists: that more of a good thing is better. This will allow me to identify sufficient conditions for a good life in line with multiple traditions, not only the one in population ethics. Let us begin our examination.

2.2 Camus’ criterion

Albert Camus is the most famous proponent of the view that “One kills oneself because life is not worth living” (Camus, 2004, p. 499). While I will refer to this view as *Camus’ criterion*, the view is

shared, occasionally with some important modifications that we will examine, also by some anglophone philosophers and other existentialists, with a tradition that can be tracked as back as the book of Job in the Bible (Benatar, 2008; Hume, 1990; King James' bible, 1998; Williams, 1995, pp. 224–232).

Camus does not add any argument to his view, as he considers it a “self-evident truism” (Camus, 2004, p. 499). Camus' statement is also imprecise. If a life is not worth living if one commits suicide, then, for a life to be barely good is for a life to be among the worst lives where one would not commit suicide. This requires us to examine what it means that one would not commit suicide.

A full exegesis of this idea is beyond the scope of this Section, but there are instructive distinctions and clarifications to be made. The first distinction is that Camus' criterion can be interpreted in a subjective and in an objective sense. The subjective sense is the following:

Camus' subjective criterion (death). A life is good iff a person living it does not *desire* to end it. A life that is barely good is the among worst good lives.

There is a weaker formulation of this idea. Some may consider ending one's life too strong. Suicide is a violent act, and some who have bad lives may not desire such violence. But there is a view that hints at the same intuition, defended for example by Bernard Williams (Williams, 1995, pp. 224–232) and expressed in the book of Job, where Job expresses his dissatisfaction with his life by asking “Why did I not die at birth, come forth from the womb and expire?” (King James' bible, 1998, v. 3.3-3.11). According to this view, that a life is bad if one does not resent it. More precisely:

The subjective criterion about birth: A life is good iff a person living it does not regret being born. A life that is barely good is the worst good life.

It is likely that these two criteria give different recommendations: it seems easier to wish never to have been born than to end one's life (Benatar has defended that this ought to be the case in his 2008, pg. 18-28). In this Chapter, I focus on Camus' criterion only, but the structure of the subjective criterion about birth is identical to Camus' criterion. All of my considerations about Camus' criterion apply to the subjective criterion about birth: the reader is free to read this Section as referring to either criterion.

The subjective interpretation of Camus' criterion (or the subjective criterion about birth) is hard to believe, as people often change their minds as to whether or not their life is worth living. There is plenty of research showing that people who tried to commit suicide and failed are unlikely to try again (Owens et al., 2002; Shaffer, 1996; Suominen et al., 2004), which suggests that they evaluate their life differently at different times.

More crucially, people wish death upon themselves for reasons that do not seem to track an evaluation of one's whole life. As Smuts puts it, "after breaking up with her latest boyfriend of two weeks, a depressed adolescent might think that her life is not worth living. Sadly, she might want to commit suicide. But this does not mean that her life is not worth living. She is mistaken. Hence, the desire to commit suicide is not sufficient for a life to be not worth living." (Smuts, 2013, p. 443). There are plenty of similar examples. Disappointments at work or boring days may make one's life feel unbearable. These considerations may not be temporary whims: on these long days, people may be convinced of the badness of their lives after critically assessing their conditions, only to feel very differently on the next day. Other factors distorting one's judgment about one's life can be mental illness, damaging environments, or damaging relationships. Since one's subjective desire for death is unreliable, Camus' subjective criterion is unpromising.

One may try to improve Camus' criterion by subtracting it from subjective arbitrariness. One may want to correct it as follows:

Camus' objective criterion. A life is good iff it is *prudentially better* for who lives it not to end it.

Remember that this discussion pertains to the *intrinsic* value of an *earthly* life, not the way in which a life can be good for others, nor whether it can be good beyond one's earthly existence. This means that whether this person would commit suicide because it benefits others, or because it gives meaning to their life, or because the afterlife appears more appealing than earthly life, are not valid reasons to identify a life that is good or bad.

However, we are left with a criterion that does not give any guidance. How does one check whether it is better to keep on living, or whether it would have been better never to have existed? Camus' objective criterion does not tell us. What does one tell the adolescent who, after breaking up with their partner, wants to commit suicide? Camus' objective criterion will struggle convincing the adolescent to keep on living, as the only way to do so is to tell them that their life is good, and the only way to tell them that their life is good, on Camus' criterion, is to insist that it is better to continue it. Any argument to persuade the adolescent to live that relies on Camus' objective criterion is circular.⁴²

We should abandon Camus' criteria, as they are either implausible or do not give guidance as to which lives are good and which are not.

2.3 Continuation Criterion

More promising may be the criterion according to which a life is good if we would like to continue living it. I call this the Continuation Criterion. This criterion is often used as a *reductio* by pessimists: as Schopenhauer puts it, "No man, if he be sincere and at the same time in possession of his faculties,

⁴² Additionally, Theron Pummer raises the point that it is unclear when is it better (not) to end a life to have a good life. That is, it is not clear, whether Camus believes that a life is bad if there is any point in time where it would be better to end it, or if a life is bad if it is better to end it at birth. I find the latter interpretation more charitable.

will ever wish to go through [their life] again” (Schopenhauer et al., 2010). Similarly, Hume has Philo saying:

Cicero [...] introduces CATO, the great, the fortunate CATO, protesting in his old age, that had he a new life in his offer, he would reject the present. Ask yourself, ask any of your acquaintance, whether they would live over again the last ten or twenty years of their life. No! but the next twenty, they say, will be better: “And from the dregs of life, hope to receive What the first sprightly running could not give.” Thus at last they find (such is the greatness of human misery, it reconciles even contradictions), that they complain at once of the shortness of life, and of its vanity and sorrow (Hume, 1990, book 10).⁴³

Similar arguments have a long tradition, with illustrious names (Bayle, 1696; Kant, 1900, p. 259; Leopardi, 1992; Nietzsche, 2006, para. 341). The sentiment at the core is that a life is good iff living the same life twice is better for a person than living it just once.

As pointed out in (King, 2022, p. 460; Leibniz, 2019, para. 13.130), the pessimist’s *reductio* dissipates, at least in great part, once we start examining what it means to live the same life *identically*. We want to avoid problems of diminishing marginal utility, that is, problems concerning goods that make one’s life less good if repeated. For example, the pleasure from a tasty cake has diminishing marginal utility: while the first bite may be pleasurable, after a certain very high number of bites we grow tired and bored of the taste, and do not want to eat the cake anymore. The cake may be good for us, despite we would not have it *identically* anymore. Of course, if we were to enjoy something else, for example a cup of tea, as exquisite as the cake, then we would have it. What we want is that the continuation of a good experience (living a good life, eating a good cake) is identical in quality, not in every feature. *Novelty* and *variation* may be needed to keep the same level of quality. Once that is

⁴³ Hume’s citation is from Cicero’s *Cato Maior de senectute*, Section XXIII. The interpretation does not seem correct to me, as I interpret Cicero as saying that no life is as good as the afterlife, so it is good to die rather than live again. But an exegesis of Cicero is beyond the scope of the thesis.

considered, the pessimist argument loses most of its bite, if not all. This means that the best interpretation of the continuation criterion is that a life is good iff continuing to live it at the same quality of life improves this life.

While the criterion does not serve the purpose that Pessimists believe, the criterion still has its merits. It is indeed true that we judge, for example, how good a cake was by asking if we would have some of it again given certain conditions (such as, for example, variation). We can define the criterion bypassing the issue of marginal utility as follows:

Continuation criterion. A life is good iff adding continuations of this life at the same average momentary quality of life would make the life better.

In introducing this criterion, I am assuming that the value of a life is an increasing function of its average wellbeing. This assumption is not clearly stated by Pessimists or previous theorists that hold the Continuation Criterion. It seems to me, however, to capture the spirit of what Continuation Criterion theorists are arguing in favour of. In fact, the core tenet is that a life is good if it would be better to live it again: any formulation of this idea that I can think of is consistent with the value of a life being an increasing function of its average wellbeing.

Does this criterion fare better than Camus'? In a sense, they are very similar: they both insist that a life is good if an addition of parts relevantly similar to what is already in the life would be better. However, the Continuation Criterion appeals to betterness in a much more straightforward way, which saves it from circularity. Consider again the adolescent who wants to die because they recently broke up with their partner. We want to tell them that they should keep on living because their recent breakup is not representative of their overall quality of life. By telling them that the continuation of their life will be good in the sense of the Continuation Criterion will indeed convey this message: this saves the Continuation Criterion from circularity.

We have seen that, in both Camus' criterion and the Continuation Criterion, a core intuition is that a life at a certain wellbeing level is good if more of this life is better.⁴⁴ I find it striking that this idea is present in both criteria, and we are just about to see that it is not alien to the one in population ethics either.

2.4 Better than neutral

The scholarship of population ethics has greatly focused on what it means for a life to be barely good, since it is crucial to evaluate the acceptability of the Repugnant Conclusion (see Chapter 5: Totalism: Between Repugnance and Greediness).

Unlike the Continuation Criterion and Camus' criterion, population ethicists do not estimate whether a life is neutral or not by the internal features of this life. Instead, they consider a life good when it is better than a "neutral level of wellbeing". The neutral level is the wellbeing level where adding a life to a population does not make the population better or worse. We have assumed that this is identical to a wellbeing level where life is neither good nor bad.⁴⁵ Sidgwick is the first to mention such neutral level, which he takes to be a necessary metaphysical truth rather than a fact of common experience:

For pain must be reckoned as the negative quantity of pleasure, to be balanced against and subtracted from the positive in estimating happiness on the whole; we must therefore

⁴⁴ Which is not to say that, for something to be good, it needs to have no marginal utility. The Continuation Criterion is supposed to apply to lives only. I thank Petronella Randell for pressing me to clarify this point.

⁴⁵ It may be worth reminding, at least in a footnote, this a distinction between what population ethicists are looking for and what Camus, Hume, Schopenhauer and the other authors I referred to in the last two Sections were looking for. All the authors in the last two Sections were looking for a prudentially good life. Population ethicists, instead, look for a life that increases population value. Some authors think that a life increases population value if and only if it is prudentially good, but not all of them: some theorists believe that the level where a life is good is different from (typically, higher than) the level where lives make a population better (Blackorby et al., 1997, 1998, 2005; Bossert et al., 2022). As stated in the introduction of this Chapter, in order to avoid the problems that these theories face (see pg 49/50), and other problems, I am assuming that a prudentially good life always increases population value.

conceive, as at least ideally possible, a point of transition in consciousness at which we pass from the positive to the negative. (Sidgwick, 1981, p. 124).

While a neutral welfare needs not to be identified in a balance of pleasure over pain, it is located between positive and negative welfare levels.

Since the neutral level is the starting point of the inquiry, the first step of population ethicists is to identify this neutral level. Once this level is identified, “A life has positive (negative) welfare if and only if it has higher (lower) welfare than a life with neutral welfare.” (Arrhenius forthcoming, pg. 24). Thus, a life that is barely good is a life that is just barely better than the neutral level.

Let us begin with the definitions closer to the Continuation Criterion. (Bader, 2022) and (Blackorby et al, 2005, Chapter 9) suggest that the zero level may be the limit that lives approach as they get shorter and shorter. That is: any good life loses value if it is sufficiently shortened, and any bad life increases its value if it gets sufficiently shortened. The neutral level must be the value where these good and bad lives converge as they are shortened.

This bears similarities to the Continuation Criterion, according to which a life is good iff it is better to keep on living it at the same quality of life. In fact, a key assumption in the definition by (Bader, 2022; Blackorby et al., 2005) is that shortening a good life makes it worse, and increasing a bad life makes it worse: this mirrors the core intuition of the Continuation Criterion, according to which it is better to live a continuation of a good life if the continuation is at the same level of quality, and it is worse to live a continuation of a bad life if the continuation is at the same level of quality. The intuition that length is a test for whether a life is good or bad remains present, despite Bader and Blackorby’s tests are about the shortening of a life, while Hume and Shopenhauer’s are about the continuation of a life.

Another popular identification of the neutral level is that it is the level that is “as good as nonexistence” (Broome, 2004, pp. 140–150; Rabinowicz, 2009a, p. 390). This account has been sometimes criticized as being incoherent. If one does not exist, then one cannot have interests, and

things cannot be better or worse for them (Adler, 2009; Bykvist, 2007; Cohen, 2020; Holtug, 2010a; M. A. Roberts, 1998, 2003). Those who believe that neutral lives are as good as nonexistence have offered several replies, for example, showing that we can meaningfully compare in betterness entities that do not exist (Arrhenius & Rabinowicz, 2010, 2015b; Broome, 2003, 2004; Buchanan & Brock, 2007; Greaves & Cusbert, 2022; Heyd, 1988, 1994; Narveson, 1967).

I will not take a stand here in this complex and important debate. All I want to say is that this account exists. And, despite being typically an account of neutrality with respect to population value, if we assume that a life is neutral with respect to population value iff it is prudentially neutral, this account has some similarities with Camus' criterion, according to which a life is good if it is not better to die, or if it is not better never to be born. In fact, to prefer death, or to prefer never to have been born, one must prefer nonexistence to existence.

And, like Camus' criterion, it is not sufficient to identify what makes a life neutral. It does not offer any guidance: to identify whether a life is as good as nonexistence we need another criterion. We will not identify whether a life is better, worse, or as good as nonexistence just by looking at it. Still, it can be useful to know that the neutral life we are looking for ought to be a good candidate for a life that is as good as nonexistence, if this comparison can be made at all.

On the last prominent account of what makes a life neutral, a life is neutral if it is, on the whole, as good as a part of a life that does not add anything of value or disvalue to one's life. As Arrhenius puts it, "A life has neutral welfare if and only if it is equally as good for the person living it as a neutral welfare component." (Arrhenius manuscript p. 33). A "neutral welfare component" is something that does not change the overall value of a life: in the words of Bykvist, "A person's life has neutral welfare at a time t if and only if her life up to t has the same welfare as her life including t " (Bykvist 2007, p. 101).

Broome endorses this definition too, but applies it to an extended life: "Call a life 'neutral' at a time if it is just on the borderline between being worth living and not worth living at that time. Call a life 'constantly neutral' if it is neutral at every time. We might say that a life as a whole is worth

living if and only if it is better than a constantly neutral life.” (Broome 2004 pg. 64). Broome seems to be stressing, again, that a life is neutral if continuing it at the average momentary wellbeing is neutral.

These were the candidates to explain what a neutral life is. Any life that is just barely better than that is a barely good life. All theories about when a life is good can be divided into two sets: the incomplete ones requiring additional principles, such as Camus’ criterion, and the criterion according to which a neutral life is as good as nonexistence, and the ones relying on some continuation criterion, according to which a life is good (bad, neutral) if a continuation of it at the same average wellbeing level would be better (worse, equally good). Since we find the first set lacking, we should accept the Continuation Criterion in this general form. Note that even Camus’ criterion seems to be an attempt to conclude that a life is good if a continuation of it is better.

2.5 A “test” for a good, bad, or neutral life: the Tradeability Principles

From the Continuation Criterion, the existence (at least in principle) of good lives, bad lives, and neutral lives, and the assumption that goods come in a smooth continuum (introduced in Chapter 2 and explored in Chapter 3) we can propose a test for when lives are good, bad, or neutral. I call it:

Positive Tradeability: L is a good life iff there is some decrease d in average wellbeing such that a life L_1 that consists of twice the duration of L, and whose average momentary wellbeing is worse by d , is a better life than L.

More formally:

Positive Tradeability (Precise): Let the value of some life L as $v(L)$. Let L_i be a life and let $W(L_i)$ be L_i ’s average momentary wellbeing over their life. Let L_j be a life with twice the

duration of L_i and average welfare $W(L_j)=W(L_i) - d$ for some positive d . L_i is a good life iff there exists some d such that $v(L_j)>v(L_i)$.

Here is how we reach Positive Tradeability. The Continuation Criterion implies

Doubling Good Lives: a life is good iff doubling its duration while maintaining the same momentary wellbeing makes it better.

I interpret the improvement that a good life gets when doubled as “non-trivial”. This means that, if we were to pay some small price in momentary wellbeing to double the life’s duration, it would still be better to do so.

Given Doubling Good Lives, L has value $v(X)$, iff a life that lasts twice as long as L at the same momentary welfare has $v(Y)$ goodness, such that $v(X)<v(Y)$. Assuming that the improvement that a good life gets when doubled amounts to some non-null improvement in value, then, for some positive amount of welfare A , $v(Y)-v(X)= v(A)$. Given the smooth continuum, there is some sufficiently small decrease of momentary welfare to L , call it d , such that $d< v(A)$. Doubling the duration of L means adding $v(A)$ to $v(X)$, and slightly decreasing the momentary welfare of that life means decreasing $v(X)+ v(A)$ by d . Call L_1 the life that results from this operation, having welfare $v(X)+ v(A)-d$. Since $d< v(A)$ and $v(X)+ v(A)>v(X)$, then $v(X)+ v(A)-d>v(X)$. Thus L_1 is better than L : QED.⁴⁶

⁴⁶ I am setting aside some difficulties. Some may worry that there are some momentary welfare levels such that, if one lives a short life at these welfare levels, this life is neutral or bad, while if one lives a sufficiently long life at these welfare levels, this life is good. As the short life would not be good but may respect Positive Tradeability, such lives would be a counterexample to Positive Tradeability. I know no argument concluding that such life exists, but I do not have an argument for denying that it does either. If such life exists, then Positive Tradeability has to be restricted to sufficiently long lives. For matters of space I will, however, ignore this complication in the rest of the thesis.

It is important to realise just how intuitively plausible Positive Tradeability is. Positive Tradeability does *not* say that a life is good if we can improve it by doubling its duration and decreasing its average momentary wellbeing by some *fixed* amount of welfare. The amount can change for each welfare level. This avoids that, by applying Positive Tradeability a sufficient number of times, we obtain the intrapersonal version of the Repugnant Conclusion. In fact, for any two welfare levels of average momentary wellbeing M and m , $M > m$, for no number of iterations of Positive Tradeability implies that a life at m is better than a life at M .

I illustrate why this is the case with what is, in effects, a version of one of Zeno's paradoxes. It follows the blueprint of (Broome, 2023). Let us represent with the natural numbers 100 and 99 two momentary welfare levels, one greater than the other. I show that no number of Positive Tradeability iterations implies that some life L1 at average momentary wellbeing 100 is worse than some life Ln at average momentary wellbeing 99, no matter how long life Ln is.

A key assumption for this argument is smooth continuum assumption we made in Chapter 2 and never abandoned. Given the continuum, we can go from any welfare level w_1 to any other welfare level w_2 in a finite number of arbitrarily small steps. Let us now assume this is valid both for lifetime wellbeing levels and momentary wellbeing levels.

Let us examine a life L1 consisting of 50 years at wellbeing level 100. Given Positive Tradeability, for some slight decrease d in average momentary wellbeing, if we double its duration and slightly decrease its average momentary wellbeing by d , we obtain a better life. Suppose that d is 0.5: Positive Tradeability says that L1, a life of 50 years at average momentary welfare level 100, is a good life because there is a better life L2 consisting of 100 years at average momentary welfare 99.5.

Now, let us iterate Positive Tradeability again on L2. It will say that there is a better life L3 that is twice as long as L2, but has slightly less momentary welfare by an amount d' , than L2. This life L3 will last 200 years, and have momentary welfare $99.5 - d'$. Since we need not to assume that

$d'=d$, we can assume that $d'=0.5^2$, that is, 0.25. So L3 is a life that lasts 200 years at a momentary welfare of 99.25.

Let us now iterate Positive Tradeability for L3. It will say that there is a better life L4 that is twice as long, but has slightly less momentary welfare by an amount d'' , than L3. Let d'' be 0.5^3 , that is 0.125. Positive Tradeability says that a life L4 of 400 years at welfare 99.125 is better than a life L3 of 200 years at welfare 99.25.

And so on: for any iteration n , we can just make d smaller and smaller (but never 0), for example by letting it be 0.5^n : this way, we will never reach welfare level 99, despite getting arbitrarily close. This way, Positive Tradeability does not commit us to the intrapersonal Repugnant Conclusion.

An argument similar to the one for Positive Tradeability can be made in favour of:

Negative Tradeability: L is a bad life iff there is some increase d in average wellbeing such that a life L1 that consists of twice the duration of L, and whose average momentary wellbeing is better by d , is a worse life than L.

More precisely:

Negative Tradeability (Precise): Let the value of some life L as $v(L)$. Let L_i be a life and let $W(L_i)$ be L_i 's average momentary wellbeing over their life. Let L_j be a life with twice the duration of L_i and average welfare $W(L_j)=W(L_i) + d$ for some positive d . L_i is a bad life if there exists some d such that $v(L_j)<v(L_i)$.

And, from Negative Tradeability and Positive Tradeability, we conclude

Suspension of Tradeability: If a life is neither good nor bad, then it is a neutral life.

The Suspension of Tradeability will be crucial in arguing in favour of the Intuition of Neutrality in Section 5 of this Chapter. We will, however, set it aside for the next Section, where we will examine some examples of “barely good lives” that have been suggested in the literature.

3. Examples of barely good life

In almost 40 years of research on population ethics, many putative examples of “lives barely worth living”, that I call “barely good lives”, have been proposed. I do not aim to report them all, as some have been contested, and others have been scarcely referred to.⁴⁷ There are, however, four paradigmatic examples of “barely good lives” in population ethics which have been often used as a benchmark as to whether the Repugnant Conclusion is acceptable. These paradigmatic examples of “barely good life” are the *Drab Life*, *Barely Conscious Lives*, the *Rollercoaster Life*, and the *Very Short Life*. Let us see them in turn.

3.1 Drab Life

The most influential characterization of a life that is barely good is Parfit’s *Drab Life* (Cowen, 1996; Parfit, 2016, p. 118; Rendall, 2015; Thomas, 2018; Venkatesh, 2020, pp. 266–267). Parfit describes it somewhat quickly: he says that, “Though there would be nothing bad in this life, the only good things would be muzak and potatoes” (2016, p. 118) This short description requires unpacking, that I’ll do in the next paragraph.

The features that this life has are (1) nothing bad is happening, and (2) nothing excellent is happening. This life is, however, good because (3) it has only “very basic sources of welfare”

⁴⁷ For example, barely good lives have been described as “not a life that differs significantly from a normal privileged life” (Ryberg, 2004, p. 239) and as “probably live lives pretty much like the lives [lived by] affluent Western people” (Tannsjo, 2002, p. 342). For a criticism of this description, see (Arrhenius, forthcoming, pg. 63-77).

(Venkatesh 2020, p. 265), some welfare component that generates low positive welfare, that is identical for the entire duration of the life. Who lives these lives are not in the position to enjoy anything else: Parfit says,

Even if we lost most external goods, some of us would have inner mental resources with which we could make our lives fairly good, by composing long poems, for example, or thinking about some intellectual problems. (...) We can suppose that lives in Drab Z would be only barely worth living, not because they would be lived by people like us who were in such deprived conditions, but because these lives would be lived by beings who would be psychologically much simpler than us. (Parfit 2016, p. 119).

Muzak and potatoes are taken to be bland, but overall pleasurable items to have in a life. This life is supposed to represent the “pig-like pleasures” that Mill considers to be worthless if compared to any amount of whatever made Socrate’s life good (Mill, 2000, p. 11).

Parfit does not tell us how long the Drab Lives are supposed to be. As he is aware, on some theories of how value is aggregated across lives, such as the theory according to which the goodness of a life is the sum of the goodness of its moments, any life that consists in only positive features can reach arbitrarily great value. Parfit clearly thinks that this theory of aggregation is wrong (Parfit, 2004, pp. 160–161), but does not give us any element to understand what he is relying on. I’ll interpret Parfit as saying that *any* Repugnant Conclusion consisting of Drab Lives of *any* duration is unacceptable.

Still, there are some issues to be addressed. As described, the Drab Life seems bad rather than barely positive. During a lifetime of muzak and potatoes, a person would be likely to get bored, and come to hate muzak and potatoes. To avoid this, we can assume that people living the Drab Life would not ever change their attitude towards muzak and potatoes, for example because they

completely forget what they ate or listened to the previous day – Parfit simply assumes that they have a psychology very different from ours.

Some are explicitly willing to accept the Repugnant Conclusion even if the Z-population is composed only of Drab Lives (Venkatesh 2020). Parfit makes it clear he would not accept such a Repugnant Conclusion (2016). As far as I know, no one who rejects some version of the Repugnant Conclusion is willing to accept a Repugnant Conclusion with this kind of lives. Thus I take it that, if a theory wants to successfully avoid the Repugnant Conclusion, it is necessary that it avoids that, for any population, there is a better Z-population consisting of only Drab Lives.

3.2 Barely Conscious Lives

Derek Parfit clearly states that Drab Lives are as good as lives of “barely conscious beings”. The idea is that, if the Z-population of the Repugnant Conclusion consists of “barely conscious beings”, then the Repugnant Conclusion is unacceptable, is a quite popular idea (Arrhenius, 2000b, p. 49, Forthcoming, pp. 63–78; Whewell, 1971; Williamson, 2021). The way these “Barely Conscious Lives” are described takes roughly two forms.

The first is the idea that “Barely Conscious Lives” are lives of insects (Arrhenius, 2000b, p. 49, Forthcoming, pp. 63–78; Sebo, 2023; Whewell, 1971; Williamson, 2021). In fact, insects seem to have a very limited capacity for welfare. Since their cognitive abilities are severely limited, they are likely prevented from achievements or relationships that would make a human life outstanding. Their nervous system (when present) is significantly more limited than ours, which presumably significantly limits the intensity of pleasure (and pain) they can experience. Their lives are short, which limits how much goodness there can be. Many people resist the idea that, for any population, no matter how numerous and no matter how well things go in this population, there is a better population consisting of sufficiently many mildly happy insects.

The other description of Barely Conscious Life is some variant of the following idea: a life is barely good if it contains just a drop of pleasure, and then nothing more of value. This can be a life

in a coma except for a drop of pleasure, or a very short life containing some mild pleasure (Arrhenius 2000, Forthcoming, Tännsjö 2002). It seems very unlikely that, for any population, no matter how numerous and no matter how well things go in this population, there is a better population consisting of sufficiently many people living in a coma except for a drop of pleasure, or living one mildly happy second.

I assume, then, that if a theory wants to successfully avoid the Repugnant Conclusion, it is necessary that it avoids that, for any population, there is a better Z-population consisting of only Barely Conscious Lives.

3.3 Rollercoaster Lives

Parfit contrasts Drab Lives with Rollercoaster Lives. Rollercoaster Lives contain anything good that you can imagine in a life, but also many bad things (Parfit 2016, p. 118). These lives may contain the greatest achievements, most meaningful relationships, and deepest pleasure one can imagine, but also its deepest horrors, anguishes, and suffering. The good things and the bad things balance one another in such a way that the goods do outweigh the bads, but just barely. Many report that a Repugnant Conclusion with Z-lives that are Rollercoaster Lives would be less repugnant, and hint or commit to perfectionist considerations (see Appendix 1) that may make it acceptable (Beard, 2020; Parfit, 2016; Thornley, 2022; Venkatesh, 2020).

The image of the “rollercoaster” suggests a life full of good and bad things, “ups and downs”, that follow one another within the course of a life. However, it has been pointed out that the goods and the bads may be differently distributed: these lives may have all the good things at the beginning and all the bad things afterwards, in what Beard and Venkatesh call “Job Lives”,⁴⁸ or they may have

⁴⁸ Jean Gové points out that the term is a bit misleading. Job’s life does start excellently, in order to then become a life of pain and suffering for a certain length of time, but after the period of pain and suffering he experiences as much goodness as he had at the start (King James’ bible, 1998, v. 42: 9-17).

all the bads at the beginning and all the goods afterwards, in what Venkatesh calls “Cinderella Lives” (Venkatesh 2020 p. 269).

This means two things. First, we need to explain why a Repugnant Conclusion with Rollercoaster Lives as Z-lives would be better than a Repugnant Conclusion with Drab Lives, or Barely Conscious Lives, as Z-Lives. Second, if one were to successfully avoid the Repugnant Conclusion when it comes to Drab Lives or Barely Conscious Lives, but not when it comes to Rollercoaster Lives, that can be considered a successful way to avoid the Repugnant Conclusion, all in all.

3.4 Very Short Life

Finally, Parfit considers lives that are just like lives in population A of the Repugnant Conclusion, full of all the excellent things that can make a life go best, except much shorter. According to Parfit, a Repugnant Conclusion consisting of such lives as Z-lives would be “significantly less repugnant” (Parfit 2016, p. 118).

For these lives to be just barely good, they need to be extremely short – in the scale of minutes, or seconds, or less. This makes it harder to appreciate these lives for our intuitions. It’s hard to understand how so much goodness can be squeezed in so little time, especially since many of the best thing in life, such as relationships, require some time to develop (Venkatesh 2020, p. 267). I expect some people to be skeptical that these lives can be just “barely good”. However, certainly, if some non-human lives were to manage squeezing all that can be good in a life in a few moments, I expect most people, like Parfit, to agree that a population of Z-people consisting of these non-human lives can indeed make the Repugnant Conclusion acceptable. Thus, I take it that, if one were to successfully avoid the Repugnant Conclusion when it comes to Drab Lives or Barely Conscious Lives, but not when it comes to Very Short Lives, that can be considered a successful way to avoid the Repugnant Conclusion.

Let us take stock. We have seen that the Repugnant Conclusion is considered unacceptable by many if the Z-lives in it are Drab Lives or Barely Conscious Lives. Instead, people are much more willing to accept a Repugnant Conclusion where Z-lives are Rollercoaster Lives or Very Short Lives. In Appendix 1. On some attempts to distinguish Drab and Barely Conscious Lives from Very Short Lives and Rollercoaster Lives I show that any explanation proposed thus far as to why we should prefer Rollercoaster Lives and Very Short Lives to Drab Lives and Barely Conscious Lives is unsatisfactory. In the next Chapter I will attempt to explain this preference myself.

In what remains of this Section I provide a positive argument for believing in the (Narrow) Intuition of Neutrality, according to which there is a range of wellbeing levels such that it is neither better nor worse to add to a population lives at these levels. In Section 2 of this Chapter I identified “positive Tradeability” as a condition for good lives, and “negative Tradeability” as a condition for bad lives. On positive Tradeability, a life is good if it is better to double its duration by slightly decreasing momentary wellbeing. On negative Tradeability, a life is bad if it is worse to double its duration by slightly increasing momentary wellbeing. In Section 3 of this Chapter I introduced some paradigmatic cases of “barely good lives”.

In the last Section of this Chapter, I show that some paradigmatic cases of “barely good lives”, namely Drab Lives and Barely Conscious Lives, do not have neither Positive nor Negative Tradeability. Additionally, lives slightly better or worse than them lack these properties, too. This means that there is a range of lives, some better than others, such that these lives are not better nor worse than nonexistence.

This confirms the (Narrow) Intuition of Neutrality, according to which there is a “neutral range” of welfare levels such that it is neither better nor worse to add a life at these levels. I show that this range is sizable, as it encompasses some lives that are thought to be barely good, but would not be good candidates for a life Z in the Repugnant Conclusion, and goes all the way down to almost bad lives. This supports Critical Range Views. But before getting to the argument in favour of the Intuition of Neutrality, we need a clearer understanding of what the “Intuition of Neutrality” is.

4. The Intuition of Neutrality

According to John Broome, “We think intuitively that adding a person to the world is very often ethically neutral. We do not think that just a single level of wellbeing is neutral[...]” (2004, p. 143). This is what he calls “Intuition of Neutrality”: there are many wellbeing levels such that adding a life at these levels does not make a population better or worse. Many philosophers, including Broome himself, as well as many non-philosophers, consider the Neutrality intuition “strongly attractive” (Broome, 2004; Narveson, 1973; Parfit, 1986, Chapter 19; Rabinowicz, 2009a, 2022b; M. A. Roberts, 2003). In axiological terms, the Intuition of Neutrality has been defined as follows:

Intuition of Neutrality (broadest): if a person is added to the world, her addition has no positive or negative value in itself. (Broome 2004 pp. 145-146)

I.e., if a person is added to the world, this addition does not increase population value. While the intuition is *prima facie* compelling as defined, many consider it too broad. In fact, it seems hard to deny that adding lives that are very bad makes a population worse, at least other things being equal (see Chapter 2, Section 4). If two populations were identical except that in one there are many people in constant agony that are absent in the other population, the population without people in agony is better. This is virtually uncontroversial in population ethics. Thus, we want a less broad definition of the Intuition of Neutrality, such as

Intuition of Neutrality (broad): if a person of *non-negative welfare* is added to the world, her addition has no positive or negative value in itself.

This intuition has many defenders in population ethics, and has been made popular under Jan Narveson's slogan: "We are in favour of making people happy, we are neutral about making happy people" (Narveson, 1973, p. 80). No doubt, this is a common position among laypeople too. This intuition is the basis to what has been sometimes called the (Alleged) Asymmetry, according to which there are moral reasons against creating people with bad lives, but no moral reason in favour of creating people with good lives (Algander, 2012; Cohen, 2020; Grill, 2017; McMahan, 1981; Parfit, 1986).

However, while still having some supporters (Boonin, 2020; Magni, 2020, 2021; M. A. Roberts, 1998, 2003, 2009; Vallentyne, 2009), this position has become increasingly unpopular amongst population ethicists, as it struggles to accommodate other intuitions. For example, the idea that extinction would be bad, at least if it were extinction by infertility, or that climate change would be bad for future generations, seem incompatible with the idea that having a happy continuation of the human species does not make things better (Boonin, 2020; Finneron-Burns, 2016).⁴⁹ Additionally, these theories endorse *Cyclical Incommensurability without Restriction to a Range*, that I showed to be irrational in Chapter 4: Transitivity of "better than" and inferences. Thus, this interpretation has to be rejected, as it is too broad still.⁵⁰

The Neutrality Intuition can, however, be made narrower. Many people have argued that "for a sizeable spectrum of wellbeing levels, adding people with lives at those levels of wellbeing does not make the world either better or worse. [...] The Intuition of Neutrality states that there is more than one neutral level of wellbeing." (Rabinowicz 2009, p. 390).

We can now define:

⁴⁹ On some attempt to avoid this implication, see (Frick, 2017). For a reply, see (Beard & Kaczmarek, 2019).

⁵⁰ Some of these theories are straightforwardly intransitive (Arrhenius 2000, Temkin 2012), so they should be rejected for the reasons we saw in Chapter 4.

Intuition of Neutrality (narrow): there is a range of wellbeing levels such that, for any life at these wellbeing levels, if such a life is added to the population of the world, its addition has no positive or negative value in itself.

Broome calls lives in this spectrum of wellbeing levels “neutral lives”, and the spectrum itself “neutral range” (2004): this has become standard terminology, and I shall adopt it here.

The narrow version of the Intuition of Neutrality is classically defended within a Critical Range framework. As we have seen in Chapter 5, Critical Range Theories present theoretical advantages, such as the avoidance of Maximal Repugnance, but are vulnerable to Greediness, and violate Benign Addition in the form of Maximal Greediness. Indeed, all Intuition of Neutrality views advanced so far in the literature, may they be Broadest, Broad, or Narrow, imply at least Maximal Greediness.

In addition to these problem, an additional theoretical challenge for Intuition of Neutrality theories is that they haven’t provided an independent, positive argument as to why we should think there is a range where lives are neither good nor bad (an exception are theories according to which there is imprecision in the betterness ranking: see Appendix 1). I provide such argument in the concluding Section of this Chapter. Then, in the next Chapter, I provide a model for Intuition of Neutrality theories to avoid any sort of Greediness.

5. Tradeability and the Intuition of Neutrality

In Section 2 of this Chapter, I examined what features a good life has, and what features a bad life has. In particular, by examining several theories about what is distinctive of good and bad lives I concluded that, at the core of each, there is what I called “Positive Tradeability” and “Negative Tradeability”. Any positive good, or any good life, has the property of “Positive Tradeability” iff a great increase in duration joined with a small decrease in momentary wellbeing makes a life better.

Conversely, any negative good, or any bad life, has the property of “Negative Tradeability” iff a great increase in duration joined with a small increase in momentary wellbeing makes a life worse.

In Section 3 of this Chapter we also examined which lives were considered candidates for a life that is barely good. In this Section we will check whether these lives have the property of Positive or Negative Tradeability. We will be particularly interested in Drab Lives, that are lives where one enjoys only very low pleasures for one’s entire life such as, for example, merely listening to muzak and eating (not particularly well cooked) potatoes. We will find that, plausibly, Drab Lives do not have neither Positive or Negative Tradeability, nor do neighbouring lives. This means that there is a range of lives that are neither good nor bad. This supports the Intuition of Neutrality, at least in its narrow version.

However, before getting to Drab Lives and Barely Conscious Lives, we will analyse whether Positive Tradeability applies to Very Short Lives and Rollercoaster Lives. If it does not, then such lives fall within the Neutral Range.

When it comes to Very Short Lives, which are lives at the quality level of A-lives of the Repugnant Conclusion but very short, things are simple: these lives would clearly greatly improve if they were longer. In fact, the only reason why these lives are *barely* good is that they are very short: by prolonging them, we certainly increase their value. Thus, Very Short Lives satisfy Positive Tradeability, and are good lives.

When it comes to Rollercoaster Lives, I must confess that my intuitions are not very clear. I seem to be relying more on logic than on perceivable differences between lives. However, this is what logic tells me. If there can be a life L with high “ups” and low “downs” such that the “ups” outweigh the “downs” only by a small margin, then surely there can be a life L- where the “ups” and the “downs” are just barely lower, but the “ups” still outweigh the “bads” by a margin that is smaller than the margin in life L, but still more than half this margin. If we just extend L- to a life twice as long and with the same momentary wellbeing, we will double the margin to which the “ups” outweigh the “downs” in L-. By doing so, we do obtain a life L1 where the positive margin between “ups” and

“downs” is greater than the positive margin between the “ups” and “downs” in L. Consequently, L1 is better than L, and Rollercoaster Lives are good lives.

However, I must confess that I struggle to reliably imagine a L1 better than L in the case of Rollercoaster Lives. We do not encounter this problem when analysing Very Short Lives. I thus accept that Rollercoaster Lives are good, but with lower confidence than for Very Short Lives.

Now, onto Drab Lives. To check whether Drab Lives possess Positive or Negative Tradeability, we need to evaluate lives of different lengths. This is because a key feature of Tradeability of lives is that lives become much better or worse if they become much longer or shorter. However, we have the intuition that living longer is living better regardless of one’s wellbeing level – we have a bias in favour of longer lives, regardless of their wellbeing level, because we think that living longer is in itself good. Some others, such as myself, are biased in favour of lives that are at least as long as a normal human life, despite knowing this is false.⁵¹ Thus, if I were to give an example where there are three lives, one lasting 30 years, one lasting 70 years, and one lasting 150 years, there is a risk that people will have already decided which life is best before I specify the wellbeing level.

When evaluating our intuitions towards lives similar to the Drab Life but with different durations, we need to eliminate that bias. I do not think I can fully eliminate that bias for everyone, but I can alleviate it. In order to do so, I will assume that continuing a life at a certain wellbeing level for any x years is as good for a person as starting a life lasting x years at that wellbeing level: while I recognise this is a substantial assumption, it has strong support in population ethics and moral philosophy more broadly (Feinberg, 1986, p. 26; Parfit, 1986, pp. 358–359; Williams, 1995, p. 228). Under this assumption, I will consider prolongation of a life rather than whole lives. In fact, the absolute difference in duration between life continuations can be much smaller than in the absolute difference in duration between whole lives – in the order of years rather than decades. Hopefully, this will decrease the duration bias.

⁵¹ A longer life is better only if it is good (or possibly neutral). If the life is bad, it is better to have it shorter rather than longer. See however footnote 46.

Suppose an international catastrophe hits, and all people in your region need to be relocated. You have three choices as to where you may be relocated but, since the catastrophe has hit everywhere, none of the choices will lead to the life as good as you currently have. In fact, these are the lives you expect to live in each country.

Country A. You will live a Drab Life, with only muzak and potatoes, for the next six years (and painlessly die at the end of the 6th year).

Country B. You will live a Drab Life, with only muzak and potatoes, for the next *three* years. However, the muzak is just barely noticeably better recorded than in Country A, and the potatoes have some traces of salt that make them just noticeably tastier than in Country A (and painlessly die at the end of the 3rd year).

Country C. You will live a Drab Life, with only muzak and potatoes, for the next *twelve* years. However, all muzak records stumble once each time they are played, and the potatoes are just noticeably staler than in Country A (and painlessly die at the end of the 12th year).

Suppose also that you don't have any preference as to whether or not to keep on living. You have already been separated from all your loved ones by the catastrophe, and in no scenario will you be in the position to continue your projects. You want to keep on living only if your life will be sufficiently good. Which country should you live in?

I do not find that there is an only one best option that is clearly the best, and that all the others are rationally impermissible. When I consider these cases, it does not seem to me that the decrease in quality is compensated by the increase in quantity, nor that the increase in quantity compensates for the increase in quality. Surely, if an option is worse with respect to quality *and* quantity than another, as country A may be, then it is going to be all things considered worse. But, if an option is better in

one respect and worse in another respect as it happens between country B and country C, when it comes to these life continuations around the Drab Life, I do not find it irrational to say that none is better than the other.

This is not how good lives behave. In good lives, Positive Tradeability holds. According to Positive Tradeability, or any life of positive momentary wellbeing level, no matter how high or low, there is a better life of slightly worse wellbeing level, but twice as long. If Drab Lives were positive, Country C would be the best option, since it consists of an improvement over the Drab Lives in Country A, which consists of an improvement over the lives in Country B. If Country C were the best option, it would not be rationally permissible to pick any other life. But it is. So Positive Tradeability does not hold for Drab Lives: Drab Lives are *not* good lives.

Nor are they negative lives. If they were negative lives, then Negative Tradeability would hold for them. According to Negative Tradeability, for any population consisting of bad lives, no matter how low their lifetime wellbeing level is, there is a worse population of sufficiently many lives with slightly higher lifetime wellbeing levels. If there were Negative Tradeability among these lives, Country B would be the best option, and it would be irrational to choose anything else. But this is not the case. These lives are not negative.

It still remains true that not all lives with wellbeing levels *at a time* like the ones of countries A to C are equally good. Surely, if the expected length of life were identical in the three countries, life in Country C would be best, in Country A second best, and in Country B worst. But these lives have different durations. In positive or negative lives, this would create a ranking from worst to best because an increase in length can compensate a decrease in quality of life. But this is not what is happening between these three life continuations.

Are these lives neutral? They are on the Suspension of Tradeability criterion, since on that criterion, a life is neutral if it is neither positive nor negative. As we have seen, some other theorists have defined a life as neutral if one would be indifferent as to whether or not to continue this life at a wellbeing level that is representative of this life's wellbeing, that is to say, at the same average

momentary wellbeing (Arrhenius manuscript p. 33; Broome 2004, pg. 64). Others have defined lives as neutral when living them is not better than not living them (Broome, 2004, pp. 140–150; Rabinowicz, 2009a, p. 390).⁵² In fact, it does not seem irrational to be indifferent between living any of these lives and not living any life. Lives in Countries A to C seem to satisfy both definitions.

I admit that my intuitions are not very clear as to whether continuing one's life in countries A to C is better than not living at all. However, I think that this is due to the bias in favour of living longer. If I knew that tomorrow I'd live one day like one of the days in countries A, B, or C, and I had the option to just skip it, I would be indifferent between whether or not to skip it. By "skipping a day" I mean being unconscious as it happens, for example by means of sleep, or coma. It is plausible, and I assume, that being indifferent between living or skipping a day is being indifferent between existing and not existing that day. Thus, I conclude, lives in countries A to C are neutral, in the sense that it is neither better nor worse to live them than not to live them. Additionally, if we are indifferent to skipping one of identical days, we are indifferent towards skipping all days. This makes these lives neutral also in the sense that one would be indifferent as to whether to live any life duration at the momentary wellbeing of countries A to C.

This indicates that there is a range of neutral lives, some better than others, but none of which is better or worse than no life at all. This is a positive argument in favour of the Intuition of Neutrality.

The Intuition of Neutrality wants the neutral range to be somewhat sizable. And indeed, my argument suggests that it is. In fact, the range certainly includes lives with wellbeing levels at as good as in countries A to C. But the range obviously extends much below lives in countries A to C. In fact, by gradually taking away welfare components from these lives, by lowering the muzak volume more and more, and by making the potatoes taste less and less, and by gradually shutting down every other experience, we cover all the spectrum that goes from the wellbeing levels of lives in Countries A to C to a life that is barely conscious, and finally to a life that is not conscious at all. This plausibly

⁵² More precisely: when adding them to a population is not better nor worse than not adding them. But I am assuming that the addition of a life makes a population better (worse, equally good) iff it is better (worse, equally good) for the individual to live this life than not to live it.

includes Barely Conscious lives, such as insect lives and life in a coma with a drop of pleasure. It plausibly includes a life in a coma as well.⁵³

Thus, we conclude, the (Narrow) Intuition of Neutrality is supported by the prominent theories of what a good life is. On the “Tradeability” criteria, informed by these theories, there is a range of neutral welfare levels that spans at least from the welfare level of a coma life to above the welfare level of a Drab Life, including the welfare level of a Barely Conscious Life. If these welfare levels are neutral, it cannot be better or worse to add lives at these levels to a population. A Repugnant Conclusion where the Z-lives are above this neutral range, as in Z-populations composed of Rollercoaster Lives or Very Short Lives, is not Repugnant on most people’s intuitions.

But how do we rank populations composed by neutral lives, some better than others? How can we capture the intuition that some lives in the neutral range are better than others, but they are all neutral? This is a matter for the next Chapter, where I introduce a new theory of population axiology, which can be interpreted as a theory between a Critical Range View and a Superiority theory. As a Critical Range View, it escapes the dilemma between Maximal Greediness and Maximal Repugnance. As a Superiority theory, this theory violates Non-Anti-Egalitarianism. Nevertheless, I argue that, in this case, this is an acceptable cost. We should accept my theory, or at least consider it among the most promising theories of population axiology available.

⁵³ There is also a case for this range to include lives that are slightly *worse than* a life in a coma, such as a life in a coma with a drop of pain (provided that the drop is sufficiently small). In order to include these lives, however, one would either need to argue that sufficiently small pains are actually neutral, or that there is something sufficiently good in prolonging a life that can compensate the minor badness of a drop pain. I find attractive the idea that some minor pain, such as hangnail pain or pinprick pain, are actually neutral but worse than unconsciousness. Some people have argued that there can be some positive value in merely existing: famously, this is Nagel’s position (Nagel, 2009, p. 2). If Nagel is right, prolonging a life can justify some drops of pain even if hangnail pain is negative. I am skeptical of Nagel’s view: for important criticism to Nagel, see (Lee, 2023). If hangnail pain is neutral or Nagel is right, then the lower bound of the range can be lower than the coma life. I consider all this a matter of future exploration, and I do not commit to any of this. I assume that the lower bound is the coma life, as it is consistent with many other accounts of neutral lives, such as (Broome, 2004, p. 259; Rabinowicz, 2009a, p. 131,133).

Chapter 7: The Structured Range View

Abstract: I propose a novel Critical Range theory, that I call the Structured Range View. The benefits of adopting this theory are that it does not imply Maximal Repugnance nor Maximal Greediness, representing a way out of the dilemma introduced in Section 5. Indeed, the Structured Range View respects all adequacy conditions except Non-Anti-Egalitarianism. However, I argue that the Structured Ranged View is justified in its rejection of Non-Anti-Egalitarianism. In addition, the Structured Range View respects the (narrow) “Intuition of Neutrality”, according to which adding or not adding lives with a range of wellbeing levels to a population does not make a population better or worse. The Intuition of Neutrality has proven hard to square with the idea that it is better to have lives with higher wellbeing rather than with lower wellbeing, but my theory respects both. I further show that, by applying the same reasoning to wellbeing components rather than lives, this ranking also explains why Rollercoaster Lives are better than Drab Lives. We should consider the Structured Range View among the most promising theories for ranking both lives and populations.

1. Introduction

We need a theory to aggregate population value. We have seen in Chapter 3 that this is a challenging task: impossibility theorems show that each theory necessarily violates some set of compelling adequacy conditions. We have seen in Chapter 5 (pp. 80/81) that the impossibility theorems can be interpreted as showing that we should accept the Repugnant Conclusion, which takes away the main objection to ‘totalist’ theories. I take this to be a sufficient reason to accept some sort of “totalist” ranking, where population value is the sum of the welfare of each individual. However, this leads to a dilemma: either “totalists” accept that there is only one single neutral level, and thus accept Maximal Repugnance, or they accept that there is a Critical Range with more than a single neutral level, and thus accept Maximal Greediness and further problems. In the previous Chapter I argued that there is a range of neutral wellbeing levels, thereby supporting the Critical Range view over views according to which there is only one neutral level.

The main goal of this Chapter is to formulate a Critical Range View that avoids the problems mentioned in Chapter 5. I call this theory Structured Range View. The Structured Range View behaves like Total Utilitarianism outside the Critical Range. Within the Critical Range, the theory is designed to respect the intuitions of the (Narrow) Intuition of Neutrality. According to the (Narrow) Intuition of Neutrality, there is a range of wellbeing levels such that, for any life at these wellbeing levels, if such a life is added to the population of the world, its addition has no positive or negative value in itself. I argue that, to better capture this intuition, we should accept that it is permissible to create a given population (of size n) iff no other available population features a set of lives (of size no greater than n) that contains greater total welfare. I call this permissibility rule the “Minimal Structure Condition”.

In axiological terms, the Minimal Structure Condition says that, within the range, a population P_1 (with a number n of people in the neutral range) is better than a population P_2 iff P_1 contains greater total welfare than the total welfare of any n people in P_2 . Two populations are equally good

if they contain the same amount of people and total welfare. If two populations are neither better nor worse than one another, nor equally good, then they are incommensurable.

However, I find it more intuitive to talk about the deontic version of this theory. I find that my intuitions are better captured by the deontic version, and the axiological features of the theory are plausibly tracked by the deontic features of the theory. Thus, the theory will be mostly described in deontic terms.

I show that this theory avoids both Maximal Greediness and Maximal Repugnance, and respects most known adequacy conditions for a theory of population ethics. It violates only Non-Anti-Egalitarianism, but I provide arguments as to why this is not a damning feature of the view, and indeed, some violation of Non-Anti-Egalitarianism may be desirable. The Structured Range View has to be considered among the most promising theories of population axiology.

Additionally, I show that a modification of my Structured Range View provides a satisfactory theory of how to aggregate wellbeing within a life. When ranking lives, this modification of the Structured Range View has all the virtues it has when ranking populations. Furthermore, it can explain why Rollercoaster Lives and Very Short Lives are better than Barely Conscious Lives and Drab Lives. This is because the goods composing Drab Lives and Barely Conscious Lives fall into the neutral range, while the goods composing Rollercoaster Lives and Very Short Lives fall outside the range.

2. The Structured Range View

2.1 Joining a Critical Range View with the Intuition of Neutrality

In this Chapter, we are looking for a theory that accomplishes three main tasks. First, we want it to be a theory that aggregates value in the correct way. I take the impossibility theorems to be a reason to conclude that the correct aggregative theory will present at least some aspects of “totalism”, that is the idea that the value of a population depends at least in part from the sum of individual wellbeing. Given the arguments in Chapter 6, we want the theory to present a neutral range. A theory with a totalist ranking and a neutral range belongs to the category of Critical Range views.

Second, we want a theory that accurately depicts the Narrow Intuition of Neutrality, according to which there is a range of wellbeing levels such that, for any life at these wellbeing levels, if such a life is added to the world, its addition has no positive or negative value in itself. In Chapter 5, I argued that standard “Critical Range” views do not adequately represent the value of lives in the range. In fact, these lives lead to the rejection of Benign Addition, under the form of Maximal Greediness, without satisfactorily avoiding Maximal Repugnance. We want to maintain that there is no requirement to create these lives without these downsides.

Third, we want this theory to avoid both Maximal Repugnance and Maximal Greediness, the two horns of the dilemma for “totalist” theories. Classical Critical Range theories imply Maximal Greediness, and are not satisfying in avoiding Maximal Repugnance. We want the range of neutral levels to behave differently from the way standard Critical Range Views treat it.

To accomplish these three tasks, the deontic theory I propose, which I call the Structured Range View, presents two parts: one with features aimed to capture the Intuition of Neutrality, the other that presents a totalist ranking, as Critical Range theories and Total Views do. The Intuition of Neutrality part of the theory is what Temkin calls an “essentially comparative” feature of the theory (Temkin, 2012, pp. 228–231, 363–456). A theory of population ethics has an essentially comparative part if the value of an outcome depends on the alternative available.

The “totalist” part of the theory, instead, refers to the intrinsic, “absolute” aspects of a population, where the value only depends on the internal features of an outcome. This is what Temkin calls an “internal aspect” feature of a theory (Temkin, 2012, pp. 228–231, 363–456).

The two parts of the theory combine as follows. There is a range of neutral wellbeing levels. Outside the range, we should rank populations as total utilitarianism does (although the theory could be compatible with other value functions, such as a prioritarian function: more on this in Chapter 8). There are *requiring reasons* to bring about the population with the greatest total wellbeing outside

the range. “A requiring reason to do an act A contributes towards making A required” (Pummer, 2023, p. 24):⁵⁴ if there is a requiring reason to A, absent other reasons against A, A is obligatory.

Within the range, there is no requiring reason to bring about any population. There are, however, requiring reasons *not to* bring about populations that present *shortfalls*. These shortfalls are identified by an essentially comparative ranking, that is the Intuition of Neutrality component of the Structured Range View. The greater the shortfall, the stronger the reason against that population.

The shortfall component of the theory presents the following features:

1. A population whose all members’ welfare is in the critical range does not fall short relative to the empty population.
2. The empty population falls short relative to populations that have individuals with welfare above the critical range and no individual with welfare below the upper limit of the neutral range.
3. A population falls short relative to another population of the same size (or smaller size, or to subsets of larger populations, where the complement is non-negative) if it has less total welfare.

Feature 1 is meant to capture the Intuition of Neutrality. It says that it is permissible, but not obligatory, to add people within the neutral range. Assume that the critical range goes from some welfare level 0 to some welfare level 10 (we will change notation in the next Section): feature 1 says that a population of n people at welfare 4 does not fall short relative to the empty population (that’s a population with no people in it), so we have *justifying reasons* to bring about this population. However, we are not required to do so, as we are in the range where no life is better or worse than nonexistence. In fact, justifying reasons to do an act A contribute towards A being permissible,

⁵⁴ For more on requiring reasons, see (Gert, 2003). More on this later in this section.

without also contributing towards A being required: if there is a justifying reason for A, absent further reasons for or against A, A is permitted but not required (Pummer, 2023, p. 24; Gert 2003).

A justifying reason to A can be stronger than a requiring reason not to A, despite not having requiring force. The requiring reason's "insufficiency to require action is not a result of its being too weak to generate a requirement" (Gert 2003, p. 9): "reasons have a justifying strength, and a requiring strength, and [...] these need not always be proportional." (Gert 2003, p. 8; See also (Gert, 2007, 2016; Snedegar, 2021; Tucker, 2023). There are reasons to add people in the neutral range whose justifying strength is proportional to the welfare of the person generated, but there are no reasons to generate people in the neutral range with any requiring strength. There are, instead, requiring reasons to add people above the neutral range, and against adding people below the neutral range. These reasons are granted by the "totalist" aspect of the theory. No reason generated from wellbeing of people in the neutral range can be stronger than the reasons generated from wellbeing of people outside the range. There are also requiring reasons not to create shortfalls within the range, but they can be outweighed by justifying reasons within the range. Many have argued that the Asymmetry, the intuition according to which there are moral reasons against creating people with bad lives, but no moral reason in favour of creating people with good lives (Algander, 2012; Cohen, 2020; Grill, 2017; McMahan, 1981; Parfit, 1986) based on the Intuition of Neutrality, may be rooted in the distinction between requiring and justifying reasons (Algander, 2012; Mogensen, 2019; Pummer, manuscript; Thomas, 2023).⁵⁵

Feature 2 is a key part of the Critical Range intuition: what is above the Critical Range is better than the empty population. Assume again, for now, that the critical range goes from welfare level 0 to welfare level 10: feature 2 says that the empty population falls short relative to a population

⁵⁵ There may be grounds to reject the distinction between justifying and requiring reasons. For some arguments against the presence of multiple kinds of reasons, see (Cullity, 2018; Rabinowicz, 2008, 2012) . Furthermore, some versions of this distinction may be susceptible to arguments similar to the one I will propose in Chapter 9. An investigation on the structure of reasons and the presence of multiple kinds of reason is important, but beyond the scope of this thesis. Those who are committed to the presence of a single kind of reason can simply think that neutral lives are evaluated on a different standard than other lives, in virtue of their neutrality.

of n people at 15. And, there are requiring reasons to create whichever population has the highest total welfare above the critical range.

Feature 2 says that it is required to create populations above the critical range, while feature 1 says that it is only justified, without there being a requirement, to create populations within the range. It may seem odd that sufficiently good outcomes are required, while insufficiently good ones are only justified, but not required.

However, there are analogous situations where this happens. Suppose you are a very prominent scientist working on plants. You are planning to go on holiday for two weeks after a semester of hard work, when at the last minute, your team informs you of what follows. Due to a very rare combination of propitious circumstances that will not repeat in your lifetime, in the next two weeks, if you cancel your holiday, you can develop the formula for a fertilizer so cheap and effective, it will permanently solve world hunger. I expect most people to agree that it is required of you to cancel your holiday in order to work on the fertilizer. In fact, by not canceling the holiday, you would fail to improve the world by such an extent, it would be a cause of moral outrage for you not to cancel the holidays.

Consider now a case identical in all respects, except that the formula you would produce if you cancel your holidays does not solve world hunger, but it still makes more efficient the quality of the farming in the region you live in, which will make food half a pound cheaper for some thousands of people. It seems that the degree of moral outrage caused if you do not cancel the holiday would be much less – indeed, we do not expect any. You would still be justified in canceling the holiday, as making food more affordable makes the world better. However, the improvement to the world is very little. Indeed so little that, it seems, you are not required to cancel your holidays.

Lives within the critical range are similar to the second case, where it was justified, but not required, to cancel the holidays: as they make little difference to the value of the world, it is permissible not to bring them into existence, although it would be justified to do so. Lives above the

critical range are similar to the first case, where it was required to cancel the holidays: as they improve the world a lot, it is required to bring them into existence, and unjustified not to do so.

There is an additional, crucial detail in this analogy: it shows that a non-required act would have a justificatory force. Suppose you promised a friend you would visit them during your holiday. It would be morally objectionable to break the promise without a justification. However, it may be justified to break such a promise in order to make food more affordable. I do not expect everyone to share this intuition – for example, some Kantians might not be willing to break their promises under any circumstances. However, if you do share my intuitions, you find that it is plausible that a non-required act, such as making food more affordable, can justify what would otherwise be an objectionable act, such as breaking a promise.⁵⁶

This intuition has an important role in the way the view I am proposing works. A non-required act can justify an otherwise objectionable act in my view, too. Specifically, the creation of people in the neutral range, which is not a required act, can justify a decrease in welfare to some already existing people in the neutral range, which is an otherwise objectionable act.

Feature 3 captures the intuition that, if we do create lives and populations, we want to create better rather than worse ones. However, to fully see which cases are captured by feature 3, we need to spell out the shortfall aspect of the theory with more precision. I proceed to do so in the next subsection 2.2. The subsection after, namely 2.3, is devoted to show the applications of the theory resulting from this “totalist” aspect and shortfall aspect, and how it respects intuitions about some key cases.

2.2 Statement of the model

⁵⁶ Similar intuitions can be found in (F. M. Kamm, 1985).

In this Section I propose a theory for ranking populations. I illustrate it as a Critical Range View based on the “narrow” version of the Intuition of Neutrality, where the Intuition of Neutrality is limited to a range. However the Broad and Broadest Intuitions of Neutrality theories (see Chapter 6) are fully compatible with this view: they just need to remove the upper bound in the neutral range, in the case of the Broad theory, or both bounds, in the case of the Broadest theory.⁵⁷ Additionally, while the relevance of the Structured Range Theory within population axiology has been motivated in the previous Chapters of this thesis, the Structured Range Theory does not depend on any argument previously made in this thesis. For example, the Structured Range Theory does not depend on my arguments about Tradeability. My theory is available to anyone who is sympathetic to the Intuition of Neutrality, that can embrace it without committing to anything explored in the previous Chapters (except the fundamental assumptions of population ethics mentioned in Chapter 2).

Since I am illustrating a Narrow Intuition of Neutrality view, I need to first define the neutral range. A life is in the neutral range only if its addition does not make a population better or worse. Given my arguments in Chapter 6, and following (Rabinowicz 2009a, 2022b, 2022c), I assume that a life is in the neutral range if it is neither better nor worse for someone to live it. However, while I think that the theory I am about to propose is more plausible given this assumption, there is no obstacle to endorse the rest of my theory if one is against this assumption.

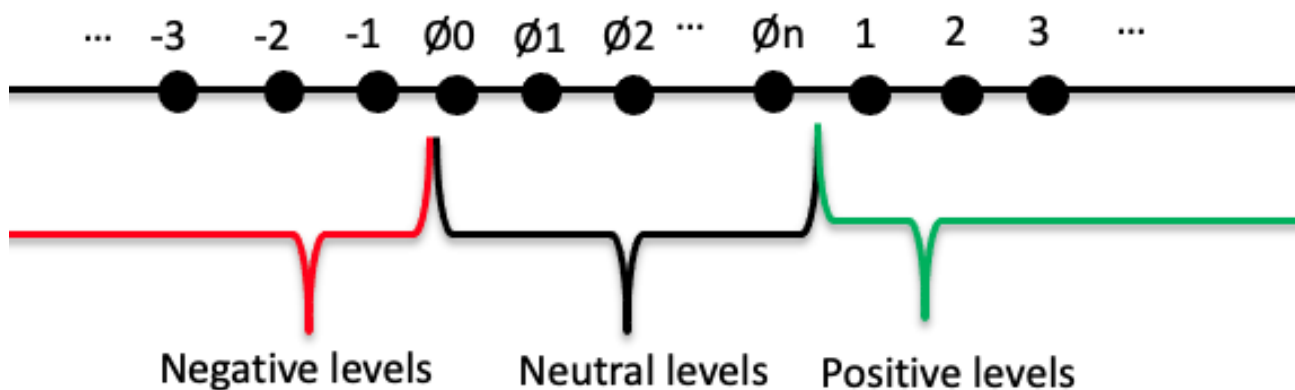
I call the lower bound of the range the welfare level \emptyset_0 (the lowest welfare level where life is neither good or bad) and the upper bound the welfare level \emptyset_n (the highest welfare level where life is neither good or bad). The neutral range is close within this interval. The welfare levels within the range can be compared within themselves as one compares positive real numbers: there are precise proportions within them, and they are on a ratio scale. However, *when comparing these levels with levels outside the range*, it is important to distinguish what is within the range and what is not, as the theory I am about to propose aggregates them very differently. The numbers outside the range can

⁵⁷ However, my theory does not help the Broadest Theory very much. The Broadest Theory will still imply Greediness and Sadism.

still be represented on a ratio scale, have precise proportions within them, are positive and negative, and we can do with them all the operations we would normally do with the ratio scale of real numbers. Still, the numbers we can represent wellbeing with are not exactly on a ratio scale, because there is no unique zero, but there is an entire range of zero levels.

Therefore, I represent what is outside the range with real numbers, and what is inside the range with real numbers preceded by the symbol \emptyset . Thus, the neutral range consists in the welfare levels between \emptyset_0 and \emptyset_n , where $(\emptyset_0 < \emptyset_1 < \emptyset_2 < \dots < \emptyset_{n-1} < \emptyset_n)$. All these numbers are neither positive nor negative, but have the same valence, that is a neutral valence. Indeed, one can interpret “ \emptyset ” as a “neutral” valence, to be distinguished from the positive valence (sometimes indicated with “+”) and the negative valence (indicated with “-”). While adding lives with positive welfare improves population value, and adding lives with negative welfare worsens population value, adding neutral lives does not improve nor worsen population value.

Therefore, I represent the spectrum of wellbeing levels as the series of real numbers, except that there is a close interval of neutral numbers between the positive and the negative values.⁵⁸ One can visualise the spectrum as follows:



⁵⁸ I am grateful to Emilia Wilson, and especially to Liam Stott, for helping me with finding the right terms to describe this structure.

To understand clearly the structure of the wellbeing levels, it is important to stress what follows: wellbeing level 1 and wellbeing level -1 are not two arbitrarily close wellbeing levels. While on a normal cardinal scale the distance between 1 and -1 is 2, here the distance between wellbeing levels -1 and 1 is 2 plus the size of the neutral range. This will have an important role in avoiding Maximal Repugnance: lives in populations Z of the Repugnant Conclusion cannot have lives that are very similar in value to lives in populations $-Z$ of the Reverse Repugnant Conclusion.

The resulting scale is one where each welfare difference can be represented as happening between welfare levels on a single scale with the structure of rational numbers, but within this overarching scale, three further scales need to be identified to talk about the absolute value of a life. One scale is for the positive values, one for the neutral values, and one for the negatives. The “zero point” in the positive value scale is its lowest point, and overlaps with the upper limit of the neutral scale. The “zero point” in the neutral scale is its lowest point, and it overlaps with the “zero point” of the negative scale. The “zero point” of the negative scale is its upper limit.

This scale is similar to the scale with which we commonly measure, for example, geographical relations such as “to the west of” or “to the east of”. Suppose we need to measure what is more to the east and more to the west among the following elements: France, Paris, Lyon, Milan, and Madrid. We will identify Madrid as the western-most element, and Milan as the eastern-most. However, when we try to estimate how far west is Madrid from France, and how far east is Milan from France, we do not necessarily take these measures from the same point. In fact, it is natural to measure how far west is Madrid from France measuring the difference in latitude between Madrid and the western-most French point. When we measure how far east is Milan from France, it is natural to measure the difference in latitude between Milan and the eastern-most French point. Likewise, when we measure how positive a wellbeing level is, I suggest that we should measure it starting from the most positive neutral level, and when we measure how negative a wellbeing level is, I suggest that we should measure it starting from the most negative neutral level. Additionally, asking whether France is eastern or western than Lyon or Paris is an ill-formed question: France contains them both. This does

not prevent Lyon from being to the more east than Paris. Additionally, the distance between Milan and Madrid is not equal to the sum of the distances of Milan and Madrid from the same point. The distance between Milan and Madrid is the distance between the south east France border and Milan, plus the distance between the south west France border and Madrid, plus the distance between these two borders.

I define the upper bound $\emptyset n$ of the range as where Positive Tradeability ends.⁵⁹ More precisely, $\emptyset n$ is a wellbeing level such that (1) if a life at this level is greatly shortened but its average momentary wellbeing slightly improved, this life would be better for who lives it and (2) if a life at this level is greatly prolonged but its momentary wellbeing slightly worsened, then this life would not be better or worse for who lives it. This is implied by Tradeability. Positive lives can be improved by increasing their duration even if one slightly decreases their wellbeing at a time. Neutral lives do not have this property, so if by shortening them and slightly increasing their welfare we improve the situation, but by prolonging it and slightly decreasing their welfare we do not, then we have found the neutral lives that lie at the upper bound of the range. While I do not have a specific example of a candidate of life living at $\emptyset n$, I can say that it has wellbeing level at a time greater than a Drab Life (see Chapter 6).⁶⁰

I define the lower bound $\emptyset 0$ of the range as where Negative Tradeability begins. More precisely, $\emptyset 0$ is a wellbeing level such that (1) if a life at this level is greatly shortened but its momentary wellbeing slightly improved, this life would not be better or worse for who lives it, and (2) if a life at this level is greatly prolonged but its momentary wellbeing worsened, then this life would be worse for who lives it. As before, this is implied by Tradeability. A candidate for a life at

⁵⁹ If anyone disagrees with me about Tradeability being a necessary and sufficient condition for good and bad lives, they can choose a different welfare level.

⁶⁰ Broad and Broadest interpretations of the Neutrality Intuitions will not have any $\emptyset n$. Broadest interpretations will have a $\emptyset 0$ that consists in either the worst logically or nomologically possible life, if there is such thing, or the life with the lowest wellbeing in the set of lives that are being considered.

$\emptyset 0$ may be a life where one is never conscious. I am open to the possibility that this life may be even lower, for reasons I explain in footnote 53.⁶¹

Now that we have a definition of the upper and lower bound, we can say something about how to aggregate lives. The aggregated value of a population outside the neutral range is the total sum of individual wellbeing levels above $\emptyset n$ minus the total sum of individual wellbeing levels below $\emptyset 0$.⁶²

The neutral range has a role in deciding which population is better than which other only as a tiebreaker, that is, when we are choosing between populations that are equally good outside the range. This is because all lives in this range are neither good nor bad for those who live them, even if some are worse than others. An improvement from something in the range to something else in the range is an improvement from (1) a life we may be indifferent as to whether or not to live, to (2) another life we may be indifferent as to whether or not to live. Conversely, an improvement from a life above the range to another life above the range is an improvement from (1) a life that is better to live than not to live, to (2) a life that is *even better* to live than not to live. The converse is true for lives below the range.

Some may find it strange that we need to prioritise improvements in some wellbeing levels, but not in others. We'll see this aspect in depth in Section 3.3 of this Chapter, when discussing Non-Anti-Egalitarianism. For now, to see why it is plausible to prioritize goods outside the range rather

⁶¹ I am also open to the possibility that there may be no wellbeing levels like $\emptyset n$ and $\emptyset 0$ in the wellbeing spectrum of actual lives, either. There may be, that is, "gaps" in the wellbeing spectrum, levels of wellbeing where no life can ever be: it may be that neutral lives can get closer and closer to good and bad lives, but no slight improvement or worsening in quality of life can take a life from neutral to good or bad. This would contradict the widespread assumption that there is a smooth spectrum of wellbeing levels (see Chapter 2. Assumptions and Methods of Population Axiology and Chapter 3 Section 4), but since it is hard to find an example of lives that are either at wellbeing level $\emptyset n$ or $\emptyset 0$, their absence from the wellbeing spectrum of actual lives cannot be ruled out. If one wants to hold that $\emptyset n$ and $\emptyset 0$ do not appear in the wellbeing spectrum, then $\emptyset n$ and $\emptyset 0$ have to be interpreted as limits: $\emptyset n$ is the limit towards which good and neutral lives converge, and $\emptyset 0$ is the limit towards which neutral and bad lives converge.

⁶² Broad Intuition of Neutrality theories calculate population value by, first, calculating the total wellbeing below $\emptyset 0$, and then using what happens with non-negative lives as a tie-breaker. Broadest Intuition of Neutrality theories are the same, but the total below $\emptyset 0$ is always 0.

than within, I ask the reader to focus on the fact that the wellbeing levels are *neutral*, and so any welfare change within the neutral level will be a change from a neutral welfare level to a welfare neutral level. It is plausible that this is a less important change than the ones from a positive wellbeing level to a better or worse one, from a negative wellbeing level to a better or worse one, from a positive wellbeing level to a negative wellbeing level, or from a negative wellbeing level to a positive wellbeing level.

It may help to visualise an example, taken from (Muñoz, 2023, p. 285). Suppose you really like hot chocolate (and believe hot chocolate is good for you), and you are indifferent between having or not having any kind of tea (and believe tea is neutral for you). However, if you do have tea, you prefer it with sugar than without. Indeed, the fact that I am indifferent between having and not having tea does not imply that I am indifferent between having sugarless tea and tea with sugar. Now, suppose you can either have hot chocolate once per month, and tea with sugar for the rest of the month, or hot chocolate twice per month, and sugarless tea for the rest of the month. It seems perfectly rational to prefer to have chocolate twice per month, and sugarless tea for the rest of the month. A possible explanation for why it is rational is that the improvement from something neutral to something better, but still neutral, such as sugarless tea to tea with sugar, is less important than the improvement from something good to something better, such as one mug of hot chocolate to two mugs of hot chocolate mug. I recognise that this example is not maximally effective, but it serves to describe the intuition I am appealing to, for now: the full argument will be delivered in Section 3.3.

We need now to decide which population to pick when populations are identical with respect to what happens outside the range but differ with respect to what happens within the range. In particular, we want to avoid the danger of Maximal Greediness. I am about to propose a model for deciding between populations that manages to do so.

The model I propose relies on the notion of what I earlier called “non-trichotomy” or violation of “completeness” (see Chapter 4), and that Broome calls a “loose ordering” (Broome, 2022) of the betterness ranking of populations. This means that, for any two populations, one of five relations may

hold. One can be better, worse, or as good as the other, as in the normal trichotomous ordering. In addition, two populations can be incommensurable, that is, none is better, worse, nor as good as the other (Chang, 2002; Rabinowicz, 2009a, 2022b; Thornley, 2022). Finally, it can be that there is no determinate answer as to whether or not one population is better than the other. As explored in Chapter 5, Critical Range theorists consider a candidate for a source of looseness in the ordering the range of wellbeing levels where a life is either incommensurable with nonexistence, or it is indeterminate whether it is better for a person to live this life or not (Broome, 2004; Gustafsson, 2020; Rabinowicz, 2009a; Thornley, 2022).

Whatever one's motivation for looseness in the betterness ordering, the looseness can only go so far. Not to prefer some populations in the range to others is irrational, even with a loose ordering. For example, if I can choose either a greater population with higher wellbeing within the range or a smaller population with lower wellbeing within the range, surely it would be irrational to pick the one that is smaller and with lower wellbeing. We need a system to represent that.

More specifically, when choosing between populations whose individuals have welfare exclusively in the neutral range, we want some structure. Chiefly, we do not want the choice of a population where everyone is worse off to be permissible. We want our theory to respect the

Minimal Structure Condition: When choosing between populations with all individuals in the neutral range, choosing a population (of size n) is permissible iff no other available population features a set of lives (of size no greater than n) that contains greater total welfare.

The Minimal Structure Condition, correctly, does not prescribe maximising the number of neutral people in a neutral population: if these people have neutral wellbeing, then any number of them is permissible. However, if people do exist in the neutral range, the Minimal Structure Condition prescribes to maximise their wellbeing. Thus, on the Structured Range View, adding people in the

neutral range is always permitted, so long as we maximise the total wellbeing of the population that ends up existing.

It may be important to note that we have to maximise the total wellbeing of the population that ends up existing even if this decreases the wellbeing of a person who would have existed regardless the addition. This way, the otherwise impermissible act of lowering a person's wellbeing in a neutral population of size n is justified by the non-required act of maximising the wellbeing of a neutral population of size $n+1$.

This concludes the description of the Structured Range View. To summarise it, we can define it as follows:

The Structured Range View: a population is better than another if it has greater total wellbeing outside the neutral range. When choosing among populations differing only with respect to people within the neutral range, we use the Minimal Structure Condition. On this Condition, choosing a population (with a number n of people in the neutral range) is permissible iff no other available population features a set of lives (with a number of people in the neutral range not greater than n) that contains greater total welfare.

I formulated the theory in terms of permissibility, as I think it is the most intuitive sense to talk about choosing between populations within the neutral range. However, I take this to be tracking the axiological structure of the neutral range. On this structure, a population $P1$ is better than another population $P2$ iff it is impermissible to choose $P2$ given the presence of $P1$ in the choice-set. If it is permissible to pick either, then the two populations are equally good or incommensurable.

More precisely, we can define:⁶³

⁶³ I thank Wlodek Rabinowicz for encouraging me to specify this formulation of the theory.

The Axiological Structured Range View: a population is better than another if it has greater total wellbeing outside the neutral range. Within the range, a population P1 (with a number n of people in the neutral range) is better than a population P2 iff P1 contains greater total welfare than the total welfare of any n people in P2. Two populations are equally good if they contain the same amount of people and total welfare. If two populations are neither better nor worse than one another, nor equally good, then they are incommensurable.

Since I find it more intuitive, I will keep referring mostly to the deontic, non-axiological formulation for the rest of the thesis.

It is worth briefly noting what follows. The theory says nothing as to how to treat incommensurability across *lives*. That is, unlike the theories listed in Appendix 1, it does not have a method to indicate when two lives are incommensurable (except for some lives in the neutral level: see Section 4.1). This is, however, not a cost for this theory, as there are different ways to incorporate ways to account incommensurability in the Structured Range View. For example by adopting a way to measure multiple values in such a way that they generate incommensurability, as in the theory defended in (Thornley, 2022). This is, however, beyond the scope of this thesis.

2.3 How to use the Structured Range View

To see how the Structured Range View works, consider the following population choice-set. These populations illustrate the attractiveness of “feature 3” illustrated in subsection 2.1 of this Chapter.

A: 1 person at $\text{Ø}9$

B: 2 people at $\text{Ø}8$

C: 3 people at $\text{Ø}4$

D: 1 person at $\text{Ø}5$

E: 1 billion people at $\emptyset 1$

Any population with total welfare outside the range greater than 0 is better than any of these populations A to E, and any population with total welfare outside the range smaller than 0 is worse than any of these population A to E.

However, if these are the only populations in the choice-set, it is permissible to choose either A, B or E, but not to choose C or D. This is because A, B and E satisfy the Minimal Structure Condition, while C and D do not. Let us see why for each population.

On the Minimal Structure Condition, when choosing between populations with all individuals in the neutral range, choosing a population (of size n) is permissible iff no other available population features a set of lives (of size no greater than n) that contains greater total welfare. Starting with population E, E has a population of one billion people with a total of one billion welfare within the neutral range. E is a permissible option because no other available population features a set of no more than a billion lives with greater total welfare than one billion within the neutral range. Thus, E satisfies the Minimal Structure Condition.

A has a population of one person and a total welfare of 9 within the neutral range. Since no other population in the choice-set features a set of no more than one life with greater total welfare than 9 within the range, A satisfies the Minimal Structure Condition and is a permissible population in this choice-set.

B has a population of two people and a total welfare of 16 within the neutral range. Since no other population in the choice-set features a set of no more than two lives with greater total welfare than 16, B satisfies the Minimal Structure Condition and is a permissible population in this choice-set. Note that a transition from A to B, on the Structured Range View, is permissible, because the otherwise impermissible act of decreasing the wellbeing of the person at $\emptyset 9$ is justified by the maximisation of the total wellbeing of the resulting population.

C is *not* a permissible population in this choice-set. C features three people, and a total welfare of 12. However, there is a population in the choice-set, namely B, that features no more than three lives, namely two, with greater welfare of 12, namely 16. Thus, C does not satisfy the Minimal Structure Condition, and is not a permissible population in this choice-set.

Similarly, D is not a permissible population in this choice-set. D features one person, and a total welfare of 5 within the range. There are two populations in the choice-set, namely A and B, that features no more than a life and have greater total welfare than D (9 and 8 respectively). Thus, D does not satisfy the Minimal Structure Condition, and is not a permissible population in this choice-set.

Some may consider it problematic that the Structured Range View makes E a permissible population to create. Specifically, some may complain in saying that this is a new instance of the Repugnant Conclusion where, for any population with average welfare on the upper range limit $\bar{\Omega}_n$, is not better than any sufficiently large population consisting only of people arbitrarily close to, but above, the lower range limit.

I do not think this is a problematic feature of the Structured Range View: lives at the upper bound of the range are *not* high quality lives. If one relies on my Tradeability definitions, illustrated in Chapter 6, these lives are indeed neutral: we may be indifferent between living any day at these levels and not living them. The difference in wellbeing between lives at the upper bound and at the lower bound is significant, but not so significant that it would be Repugnant. While in Chapter 6 I did not commit to any life being the upper and the lower limit of the neutral range, all that matters for the size of the range to be appropriate is that it includes Drab and Barely Conscious Lives while excluding Rollercoaster and Very Short Lives. Thus, we can say that the upper bound $\bar{\Omega}_n$ is at least as good as a slightly better than a Drab Life, and the lower bound $\bar{\Omega}_0$ is not better than a life where one is constantly in unconscious.

This means that the closest that the Structured Range View gets to the Repugnant Conclusion may be the following. For any population consisting of lives that are slightly better than Drab Lives (say, they have a drop of salt in their potatoes and their muzak is better recorded than in Drab Lives),

there is a not worse population consisting only of lives that are spent unconsciousness, except for an occasional drop of welfare. I do not think this is a disturbing version of the Repugnant Conclusion.

Some may want for the range to be wider. Namely, some will think that the upper bound must be higher than “slightly better than Drab Lives”, or that the lower bound should be lower than a life of unconsciousness. Nothing I say in this thesis prevents that. However, the wider the range is, the worse the Repugnant Conclusion-like implication that the Structured Range View gets between populations like E and a comparatively smaller population at \emptyset_n becomes. In general, the wider the range is, the more the wellbeing levels at the extremes of the range will look different in kind, and it will be implausible that they are all neutral. Thus, I do not commit to more than the upper bound of the neutral range being slightly better than Drab Lives, and that the lower bound is a life of consciousness. This makes acceptable the fact that, for any population consisting exclusively of people close to the upper bound, there is an equally choice worthy population consisting of people close to the lower bound. And, it makes for an acceptable avoidance of Maximal Repugnance, for reasons explored in Chapter 6.

Let us now move on, and add a couple of more options in this choice-set, to see some interesting features of the Structured Range View. Namely, I add populations that allow us to better test how this model captures the neutrality intuition: the empty population, and a population like A, but with an additional person at \emptyset_9 . As a bonus, we will encounter an interesting feature of my theory: that whether or not a population is permissible in my model depends on what other populations are in the choice-set.

Consider the following population choice-set, identical to the previous except for population F and G:

A: 1 person at \emptyset_9

B: 2 people at \emptyset_8

C: 3 people at \emptyset_4

D: 1 person at $\emptyset 5$

E: 1 billion people at $\emptyset 1$

F: 2 people at $\emptyset 9$, 1 person at $\emptyset 3$

G: no one

In this choice-set, populations that are permissible are A, E, F, and G. Populations that are *not* permissible are B, C, and D. A and E are permissible for the same reasons as in the earlier population choice-set, and C and D are impermissible for the same reasons as in the earlier population choice-set, and I will not repeat them here.

However note that, while population B was permissible in the previous choice-set, B is now impermissible. In the previous choice-set no other population in the choice-set features a set of no more than two lives with greater total welfare in the neutral range than 16. This satisfied the Minimal Structure Condition. In this second choice-set, population F has a sub-population of two people with 18 total wellbeing in the neutral range, so B does not satisfy the Minimal Structure Condition. Thus, population B is not permissible in this choice-set. This makes the Minimal Structure Condition *choice-set dependent*: whether or not a population is permissible depends on what alternatives are in the choice-set.

Population F, instead, is permissible: no other population has a sub-population of at least three people at more than 19 total wellbeing within the neutral range. The contrast between F and B is instructive, in that it shows how condition (2) is *not* a condition about the average wellbeing of people: B has higher average wellbeing than F (B=8, F= 6.3 in the neutral range), but F rules out the permissibility of B.

F is interesting for another reason if compared to A. F consists in A, plus the addition of a couple of people. Both A and F are permissible, due to the Minimal Structure Condition. This captures the Intuition of Neutrality: it is permissible to either add more people within the neutral range, or not to add these.

This should become even clearer when we consider that population G is permissible. No population has a sub-population of the same size as G with a greater total wellbeing. G 's size is 0 , and G 's total wellbeing is $(\emptyset)0$. In fact, no population in *any* choice-set can *ever* be a subset of 0 people with total wellbeing greater than $(\emptyset)0$. This means that, in any choice-set of populations exclusively in the neutral range, the empty population will always be permissible. Again, this captures well the Intuition of Neutrality: it is permissible to either add more people within the neutral range, or not to add these.

As it is, the ranking has many of the advantages of critical level rankings, plus the advantage of capturing the Intuition of Neutrality. Let us see how it fares with respect to the desiderata of any plausible population axiology.

3. The challenges of population ethics and the Structured Range View

Now that we have stated the Structured Range View, we can examine how it fares with respect to the various adequacy conditions for any theory of population ethics we have mentioned so far. I go through them in the same order I examined them throughout the Chapters of this thesis: I will examine in 3.1 how the Structured Range View deals with Benign Addition and related issues, in 3.3 how it deals with Non-Anti-Egalitarianism, in 3.4 how it deals with the objections concerning Transitivity, and in 3.5 how it deals with the Repugnant Conclusion and Greediness, including their maximal versions.

In 3.2 I analyse an oddity that the Structured Range Theory has, and has not been recognised in any other axiology: when planning one's future choices in advance, according to the Structured Range Theory there may be options that are impermissible at the time of planning, but become permissible at a later time. However, I show that this is not an axiological problem, but rather a widespread problem when facing sequential choices; furthermore, there is a plausible solution to this problem, close in spirit to the Minimal Structure Condition.

3.1 Respecting Benign Addition, Non-Sadism, and Separability

The first adequacy condition we consider is:

Benign Addition. For any two populations P_1 and P_2 , if the wellbeing level of everyone in P_1 is higher in P_2 , and there are additional people with a non-negative wellbeing level, then P_2 is better.

To see why the Structured Range View respects this condition, remember that the theory behaves like Total Utilitarianism outside the neutral range, and applies the Intuition of Neutrality within the range through the Minimal Structure Condition. I show that Benign Addition is respected by Structured Range View (1) when the welfare of every individual of all populations involved is above the neutral range and (2) when the welfare of every individual of all populations involved is within the neutral range. We do not need to check for cases where some people are within the welfare and some are above, since considerations for what happens above the range always trump what happens within: if the Structured Range View applies in case (1), then it applies also when the welfare of people involved is partly within the range, partly above. There are no other cases where Benign Addition may apply.⁶⁴

The Structured Range View respects Benign Addition when the welfare of every individual of all populations involved is above the neutral range because the Structured Range View behaves like a Total View above the range. On a Total View, a population is better than another if it has greater total wellbeing. For any population P_1 , an increase in the wellbeing of people who exist in that population and an addition of further people of positive wellbeing, necessarily, results in a population

⁶⁴ Note however that the Structured Range View does not respect Benign Addition if P_1 is the empty population and P_2 has all individuals in the neutral range. I do not think this is a problematic violation of Benign Addition, that is more plausible when comparing non-empty populations.

P_2 with greater total wellbeing than P_1 . Thus, on a Total View, P_2 is better than P_1 , and so it is for the Structured Range View so long as P_1 and P_2 are above the range.

Furthermore, the Minimal Structure Condition guarantees that the Structured Range View respects Benign Addition when the welfare of every individual of all populations involved is within the neutral range. Benign Addition prescribes that P_2 is better than P_1 , where P_2 consists in a population with the same size as P_1 , but with higher wellbeing, plus other people with non-negative wellbeing. On the Minimal Structure Condition, when choosing between populations with all individuals in the neutral range, choosing P_1 (of size n) is permissible iff no other available population features a set of lives (of size no greater than n) that contains greater total welfare than P_1 . Benign Addition makes P_2 a population of the required size n and greater total welfare than P_1 , so the Minimal Structure Condition makes it impermissible to pick P_1 over P_2 : Benign Addition is satisfied.

It is at least as easy to explain why the Structured Range View respects

Non-Sadism. It cannot be better to add bad lives than good lives to a population.

In the Structured Range View, all bad lives have negative contributory value, all neutral lives have neutral value, and all good lives have positive value: this makes it impossible to reach any kind of Sadism.⁶⁵

The next adequacy condition we consider is:

⁶⁵ If we accept that lives in a coma except for a drop of pain may be neutral, as suggested in footnote 53, some may object that there are some lives that have only negative features that, in the neutral level, may outweigh other lives. But, as I said, the condition for accepting that these lives are neutral lives is to accept that either the drop of pain is not bad or that longevity can justify it.

Separability of lives. The goodness of conferring some benefit on one person, or of bringing some people into existence, should not depend on how many other people enjoy that benefit or already exist—e.g., on distant planets. (adapted from (Nebel, 2022, p. 11). See also (Blackorby et al., 2005, sec. 5.1.; Broome, 2004, pp. 117–131; Mulgan, 2001; Parfit, 1986, p. 420; Thomas, 2022b))

It is easy to see how the Structured Range View respects Separability of Lives. All aspects of the Structured Range View rely on some ranking that compares the total sum of wellbeing of the individuals in a population, be it “totalism” outside the range, or the Minimal Structure Condition. If the ranking is the total sum of wellbeing, the contributory value of an increase in welfare, or of bringing a person into existence, to the total will be identical, no matter how many people already exist or are at that welfare level. That is, if the value of a population is the sum of the welfare of individuals, for any welfare x and any welfare y , the contributory value of x in the sum $x+y$ does not depend on y . Thus, separability is satisfied.

3.2 Sequential Choice⁶⁶

While the Structured Range View respects Separability, it still results in strange interactions between past choices and future choices. More specifically, the Structured Range View features problems with sequential choice: when planning one’s future choices, there may be options that are impermissible at the time of planning, but become permissible at a later time, or vice versa. It is not chiefly an axiological problem, but a problem within decision making. And, while the Structured Range View may be more prone to sequential choices than other views, I suspect any axiology will face problems with sequential choices: they are a general problem with decision making. (I will not argue here for the problem being general. For an overview on the problem, see Buchak 2017, pp. 175-178). Still, it

⁶⁶ I am grateful to Elliot Thornley, Pietro Cibinel, Daniel Ramöller and Matt Clark for precious discussions about this Section.

is important to point out the oddities of my theory to fully assess its plausibility. And, I will point out, there is at least one plausible way out of the oddity, although its full exploration will not be covered in this thesis.

The argument that follows, showing that the Structured Range View (and other Intuition of Neutrality views) are subject to problems with sequential choice, can be found in (Thornley, Unpublished). Suppose you are given two populations you may choose to create – for example, how to plan your family. No one but the future people will be affected.

Population A: 1 person at $\emptyset 9$

Population B: 2 people at $\emptyset 8$

Population C: 1 person at $\emptyset 8$

According to my model, either Population A or Population B is a permissible choice, as they both satisfy the Minimal Structure Condition. Only C is *impermissible*. Let us suppose that you pick B, and suppose that the two people will be born two years apart.

Fast forward at least nine months, the first person, and only the first person, is now born. You are now in Population C: there is one person at welfare level $\emptyset 8$. Population A is not available anymore, as you decided to conceive a child at $\emptyset 8$ rather than at $\emptyset 9$. Instead, Population B is still available. This means that your choice-set is now:

Move to population B: 2 people at $\emptyset 8$

Stay at population C: 1 person at $\emptyset 8$

With *this* choice-set, *without* population A, either option is permissible. C and B satisfy the Minimal Structure Condition in this choice-set. But, if we pick C, we obtain an outcome that was impermissible

when A was available. Thus, in the process of creating the permissible Population B, we transformed Population C from impermissible to permissible. I expect some to consider this a problem.⁶⁷

The problem, more than axiological, is in the decision procedure: simply comparing sets of options at each stage leads to, first, characterizing an option as impermissible, and then making it permissible. This problem is far from being exclusive to the Structured Range View. In fact, it is a general problem with sequential decision making. To see why, consider the case of a smoker who is an expected utility maximiser. The smoker knows that smoking many cigarettes dramatically increases the risk of lung cancer. Given this, when planning their future choices in advance, the smoker will want to smoke fewer rather than more cigarettes. However, the smoker also knows that they will enjoy each cigarette, and each individual cigarette makes little to no difference as to how likely lung cancer is for them. Thus, *each time* the smoker is considering lighting a cigarette, it is better for the smoker to do so, as it will bring them enjoyment without making a relevant difference as to the smoker's risk of cancer.

The smoker has a problem of Sequential Choice: the preference *rational over time* to smoke fewer rather than more cigarettes is inconsistent with the preference *rational at each time* to always smoke an additional cigarette. The smoker is not assuming the Structured Range View to derive this problem of sequential choice: this problem arises for any theory according to which it is rational to do what is best in expectation. There is a flourishing literature studying sequential choices, focusing

⁶⁷ Note that this sequential choice between A, B and C can be “money pumped” (Gustafsson, 2022a; Gustafsson & Rabinowicz, 2020; Rabinowicz, 2000). Suppose that Anna is a person who believes in the Structured Range View initially chooses A. As they have decided, a tempting demon tells Anna that, if Anna switches from A to B, the demon will give Anna a penny. Since B is not worse than A, and it is better to have an additional penny than not to have it, it would be rational for Anna to switch from A to B and get the penny. Then, when the first child at Ø8 exists, the demon offers Anna a penny to choose C rather than B: again, since C is not worse than B, and it is better to have additional penny than not to have it. Once Anna is set on C and has received the money, the demon can ask her to pay three pence to change reality is such that option A is still available. Losing three pence is worth giving Anna's children a better life, so Anna should pay the money. The demon can then re-start the offers, getting infinite money out of Anna. The solution out of this “money pump” is the same solution I propose to sequential choices: a “resolute approach”. I will discuss this issue more in Section 3.4.

on cases like the smoker, procrastination, and climate change (Ainslie, 2001; Andreou, 2006, 2007; Bratman, 2012; Gustafsson, 2022a; Gustafsson & Rabinowicz, 2020; Rabinowicz, 2000). This literature makes no axiological assumptions about populations: the problem of sequential choices is widespread.⁶⁸

To avoid this problem, we may adopt a decision procedure according to which, when planning our future, we ensure that we stick to the original plan even if, as the sequence unfolds, options diverging from the plan would be rational. This is what happens in *Resolute approaches* (Buchak, 2017; McClennen, 2012; Rabinowicz, 1995). Among these strategies, there are multiple candidates allowing us to formulate a decision procedure where Population C remains *impermissible* even after Population A is not available among the outcomes we can realise, which is intuitively the correct result.

I find that a very natural decision procedure to implement into the Structured Range View is the following Resolute Approach to choose among neutral populations:

Resolute Structure (in a choice sequence): When deciding among permissible populations consisting exclusively of neutral lives, once we choose a population P_1 of any size n at any choice-point c of a sequence of choices, we can, at any choice-point later than c of a sequence of choices, switch for another population P_2 only if P_2 features a set of lives (of size no greater than n) that contains at least as much total welfare as P_1 .

⁶⁸ Some may complain that there is a relevant difference between the smoker, procrastination, and climate change, that get problems of sequential choices during a long sequences of choices, and the Structured Range View, that get problems with only three choices. This may be a difference, but it is hard to see why it is a relevant one, and crucially, there are known problems of sequential choices involving only three options. The classic case is the one of Odysseus, who has to decide how to sail past an island populated by sirens, whose singing creates the desire to stay forever with them on the island. Odysseus knows what follows. Before listening to the sirens, he will prefer to sail past than to stay. But as he is sailing near them, he will prefer to stay.

Resolute Structure means that, in a sequence of choices, any time we choose a population of size n in the sequence, we can switch that population only with populations where there is a subset of at least n people with at least as much total welfare. In other words, on Resolute Structure, once we have chosen a population P_1 with a certain size, we can switch for a population P_2 smaller than P_1 only if P_2 has at least as much total welfare as P_1 .

This should rule out the problem of sequential choice, at least in Thornley's case: we can never pick C over B . In the first choice-set, the one including A , B and C , A rules B out. In the subsequent choice-set, the one with only B and C , we cannot pick C since, having chosen B at an earlier stage, we can only switch for populations having no more than two people whose total wellbeing in the range is at least as great as 16.

There may be better theories than the Resolute Structure for our purposes. Evaluating theories for sequential procedures, and assessing which such procedure can best be implemented on the Structured Range View, is an important issue I do not have the space to explore in this thesis. However, there is at least one approach, namely the Resolute Structure, that is a promising way out of the problems I highlighted in this Section.

3.3 The Only Acceptable Violation of Non-Anti-Egalitarianism

In this subsection we explore how the Structured Range View deals with Non-Anti-Egalitarianism, Hypersensitivity, and Hyposensitivity. I show it does not respect any of them.⁶⁹ However, it does so in a way that we should not consider damning for my theory.

Arrhenius (2000) proposes as an adequacy condition for any theory of population ethics:

⁶⁹ Since Hypersensitivity and Hyposensitivity are not tightly defined (how much sensitivity is too sensitive?), it is not clear whether the Structured Range View actually does not respect them. I steel man the position against the Structured Range View, and interpret it as violating them both.

Non-Anti-Egalitarianism. For any two populations P_1 and P_2 , if P_2 has higher average wellbeing than P_1 , higher total wellbeing than P_2 , and perfect equality, then P_2 is better than P_1 .

Violating Non-Anti-Egalitarianism is often realized by adding cut-off points to the spectrum of goods that one can experience: above the cut-off points there are Superior goods, and below there are Inferior goods. A theory violating Non-Anti-Egalitarianism typically either gives too much weight to the transition across the cut-off point, and is considered *Hypersensitive* to the cut-off point, or it gives too little weight to what happens below the cut-off point, and is considered *Hyposensitive* to the cut-off point.

The Structured Range View certainly violates Non-Anti-Egalitarianism. Here is why. The Structured Range View says that, if we have arbitrarily many people at \emptyset_n , and we can get one single person to +1 above the range by lowering everyone else's wellbeing from \emptyset_n to \emptyset_1 , it is better to do so. This change is a loss in total utility, however, on the Structured Range View, what is above \emptyset_n contributes to population value, whereas what is between \emptyset_n and \emptyset_0 is only a tiebreaker. My theory, one may argue, is hyposensitive at the upper neutral level threshold, as it gives too little weight to what happens within the range if compared to what happens outside the range.

I agree that this implication is an intellectual cost to pay, and I would like to avoid paying this cost if I could (although I think it has desirable practical applications: see page 163). It is surprising to me that a gain or loss with respect to wellbeing is more important at some wellbeing levels rather than others, even if the loss is of the same magnitude. I find it very plausible that a wellbeing increase of some given magnitude from any two wellbeing levels should be treated identically. Indeed, in the next Chapter, I will criticize Prioritarianism on a similar basis, and in Chapter 9 I will criticize the assumption of Irreducibility that Pluralists have precisely on the basis that it is hyposensitive. Respecting Non-Anti-Egalitarianism, and avoiding both Hypersensitivity and Hyposensitivity is not an impossible feature to achieve for a theory: it is one of the virtues of total utilitarianism. However,

I could not avoid this implication in my theory. I should provide a convincing justification as to why lives in the neutral range are worth so little if compared to others. I will provide two.

3.3.1 The best Superiority view

The *first justification* is that no amount of neutral items can be better than any amount of good items. This seems to me self-evident. If there is a neutral range of welfare levels, as I argued in Chapter 6, then no amount of people whose welfare is neutral can be better than any amount of people whose welfare is positive.

In Section 4 of Chapter 3 we have seen that the plausibility of Non-Anti-Egalitarianism as an adequacy condition depends on how implausible theories about Superior and Inferior Goods are. According to these theories, any amount of Superior Goods is better than any amount of Inferior Goods. These theories face a challenge: to identify in the wellbeing spectrum a sharp threshold below which wellbeing levels are Inferior, and above which they are Superior.

Since most theorists have looked for this threshold among positive wellbeing levels, no plausible candidate for this threshold has been found. It is implausible that at a certain point there are two goods, one better than the other by an arbitrarily small amount, such that one is Inferior and the other is Superior.

However, if the threshold is the one dividing neutral from positive wellbeing levels, the distinction becomes indubitable. We know that there are positive goods, and that there are negative goods. If there is a smooth continuum between positive and negative goods, necessarily, at a certain point in the continuum there will be some pair of goods, one better than the other by an arbitrarily small amount, such that one of the goods is positive, and the one after it will not be positive, or it will be vague whether or not it is positive. Eventually, down along the spectrum, at a certain point there will be some pair of goods, one worse than the other by an arbitrarily small amount, such that one of

the goods is negative, and the one after it will not be negative, or it will be vague whether or not it is negative.

While it is certainly surprising and interesting that the transition between a positive (or negative) good and a neutral good may happen over a trivial difference, and there is an important unresolved puzzle here as to where this transition happens, *it is not controversial that it happens*. And, we do not need even need to specify a plausible candidate for the threshold where the transition happens (although Tradeability provides us with sufficient conditions for that: see Chapter 6: Barely good lives and the Intuition of Neutrality) in order to argue that neutral welfare levels are Inferior to the Superior positive levels: this is immediately evident. If it is uncontroversial that there are positive and neutral welfare levels, and it is uncontroversial that neutral levels are Inferior to positive levels, then the presence of a neutral range directly implies a violation of Non-Anti-Egalitarianism.

But I need much less in order to reject Non-Anti-Egalitarianism (or justify Hypersensitivity and Hyposensitivity). The plausibility of these conditions depends on the absence of a plausible candidate for Inferior and Superior Goods. The distinction between neutral and positive wellbeing levels is have a plausible candidate for Inferior and Superior Goods: the condition for making Non-Anti-Egalitarianism (or the avoidance of Hypersensitivity and Hyposensitivity) an adequacy condition for any theory of population ethics is undermined.

There is, however, an advantage that standard Superiority views have, and the Structured Range View does not have. Superiority views say that Inferior goods cannot (always) compensate for Superior goods: the rejection of Non-Anti-Egalitarianism is explained by having different kinds of goods, some more important than others. The Structured Range View does not differentiate between different kinds of goods (Superior and Inferior), but between different kinds of wellbeing levels (Neutral and non-neutral). On the Structured Range View, some wellbeing levels are more important than others, without there being differentiations between different kinds of goods.

This makes the Structured Range View vulnerable to an objection unavailable against Superiority views. On the Structured Range View, it is more important to prevent a decrease in one

drop of pleasure in a person with positive wellbeing level than to give several drops of pleasure to many people within the neutral range, so long as the people in the neutral range would not access positive wellbeing if they were to gain this pleasure. This is a problematic outcome: how can it be better to have less total pleasure drops than more total pleasure drops? Superiority views avoid this outcome, as each drop of pleasure belongs to the same kind of good, so they should all be evaluated equally.

However, I see no reason to think that what matters is the amount of wellbeing units in a population rather than how well off each individual is. If the wellbeing units compose live's wellbeing in such a way that people living these lives would be indifferent as to whether or not to live, then it is not absurd to think that this wellbeing should matter less than the wellbeing of lives that it is better to live, or worse not to live. An improvement from a state one is indifferent about to another state one is indifferent about may well be less important than an improvement from a good state to a better good state, or from a bad state to a better bad state. While there is no reason to think that goodness depends on wellbeing units rather than individual wellbeing, there are reasons to reject the difference between Superior and Inferior goods (see Chapter 3, Section 4). Thus, we should prefer the Structured Range Theory to Superiority views.

While the violation of Non-Anti-Egalitarianism is an intellectual cost to pay, I think it is worth the advantages of the Structured Range View. In addition, by considering the practical applications of Non-Anti-Egalitarianism, I think that violating Non-Anti-Egalitarianism may be desirable, especially around the neutral level. This is my *second justification* of the Structured Range View.

3.3.2 An argument against Non-Anti-Egalitarianism around the neutral level

Suppose, as plausible, that the happiest insect lives have still very low wellbeing level – indeed, neutral levels. The number of insects is estimated around 10^{12} : that's a billion time more than the number of humans. For any unit of most basic resources that humans can benefit from (eg. Food,

heating), there are many insects that would each benefit much more than the human does. For example, a single loaf of bread may be a single meal for a person, while it is likely to be multiple meals for a colony of ants. Thus, if we were to transfer resources from humans to insects, we would most likely be increasing the total and average of wellbeing units in the world, and decreasing inequality between us and insects.

Given the great difference in numbers between humans and insects, Non-Anti-Egalitarianism would prescribe to transfer virtually all human resources to insects. I find this hard to believe. The only ways to avoid this conclusion are either assuming an implausible form of speciesism, or assume that there are Inferior and Superior goods, which I have argued against in Chapter 3, or assume that there are some changes in wellbeing levels, namely the ones happening within the neutral range of wellbeing levels, that matter less than others. We should accept this latter conclusion, which is what the Structured Range View prescribes. While the violation of Non-Anti-Egalitarianism is a bullet to bite, we should be willing to bite it – after all, if we don't do so, our theory of population ethics would prescribe to transfer virtually all human resources to insects.

It is important to make a clarification. My argument does not imply that, given that neutral welfare levels are Inferior to positive welfare levels, we can violate Non-Anti-Egalitarianism between two positive wellbeing levels, or we can embrace Hypersensitivity or Hyposensitivity at any cut-off point. Non-Anti-Egalitarianism *remains* an adequacy condition if its violation happens between welfare levels with the same absolute value (two negative levels, two neutral levels, two positive levels), until there is a strong argument to believe the contrary. And, we should be unwilling to accept any Hypersensitivity or Hyposensitivity that does not involve a comparison between neutral levels and non-negative levels. As seen in Chapter 3. The Benign Addition Paradox despite a great support towards Superiority theories, no such argument has been found.

This means that Hypersensitivity, Hyposensitivity, and a violation of Non-Anti-Egalitarianism are to be considered fatal to a theory unless it happens between welfare levels of different valence (i.e. not both negative, neutral, or positive) as instantiated in the Structured Range

View. In the next two Chapters we will see that two theories, namely Prioritarianism and Value Pluralism, are Hypersensitive or Hyposensitive: these theories have no justification for this violation nearly as strong as the Structured Range View, and I will indicate this as a reason to reject them.

To conclude, violating Non-Anti-Egalitarianism is an undesirable feature of the Structured Range View, but not a fatal one. It is not so unacceptable that an improvement between wellbeing levels that are neutral is worth less than an improvement between wellbeing levels that are positive or negative. Indeed, it may even be desirable, as accepting Non-Anti-Egalitarianism implies that we should transfer all of our resources to insects. We should consider a violation of Non-Anti-Egalitarianism a sufficient reason to reject a theory only if it does not happen without a justification as strong as the fact that nothing neutral is better than something good.

I do not expect that the Structured Range View will necessarily convince those who are committed to accepting any version of the Repugnant Conclusion, even at the cost of Maximal Repugnance, including the one involving coma lives with only a drop of pleasure or of pain. However, even those who are willing to accept Maximal Repugnance agree that is it a cost for their theory (see the concluding Section of Chapter 5: Totalism: Between Repugnance and Greediness). While I personally find my violation of Non-Anti-Egalitarianism more convincing than Maximal Repugnance, I find the opposite intuition completely justified. I expect those who embrace Maximal Repugnance to recognise the Structured Range View as the most promising non-Repugnant alternative.

3.4 Transitivity

The next adequacy condition for any theory of population ethics is:

Transitivity of “better than”: If P_1 is better than P_2 and P_2 is better than P_3 , then P_1 is better than P_3 .

It is easy to see how this adequacy condition is satisfied by the Structured Range View outside the neutral range. The Structured Range View uses a “totalist” ordering to rank populations outside the range, and the Minimal Structure Condition within the range, which requires to pick a population size whose wellbeing we should maximise, which is another kind of “totalist” ordering to choose between populations. On any theory using either kind of totalist ordering to choose between populations, a population is better than another if it has more total wellbeing. Since the relation “to have more total wellbeing than” is transitive, any totalist ranking (including the kind of totalist ranking that the Minimal Structure Condition implies) is transitive.

What needs an investigation is whether the choice-set dependence of the Minimal Structure Condition violates transitivity of “more choice-worthy than”, and “better than” in its axiological version.⁷⁰ It does so, but not in a way that is susceptible to any problem such as the ones concerning impartial information, such as the *Fickler Kidnapper*, that we saw in Chapter 4.

To see why the Minimal Structure Condition violates transitivity of “better than” (or “more choice-worthy than”) in the following way. Consider the following populations.

A: 3, 3, 3, 3

B: 5, 1, 1, 1

C: 4

A is more choice-worthy than B, since A has the greatest total wellbeing when four people are involved (12 for A, 8 for B). B is more choice-worthy than C, because it has greatest total wellbeing

⁷⁰ I want to thank Wlodek Rabinowicz for precious inputs on this issue.

when a single individual is involved (5 for B, 4 for C). However, it is permissible to pick either A and C (they are incommensurable). This violates transitivity of “more choice-worthy than”, that would conclude that A is better than C.

However, the theory is “acyclic”, which avoids problems of partial information as in the *Fickle Kidnapper* case. This means that, for any populations P1, P2 and P3, if P1 is more choice-worthy than P2, and P2 is more choice-worthy than P3, then P3 is not more choice-worthy than P1. So if we were deciding between any three populations X, Y, and Z, and we knew that X is better (or more choice-worthy) than Y, and Y is better (or more choice-worthy) than Z, then we can rationally pick X, as Z will be no better (or not more choice-worthy) than it.

To see that the Minimal Structure Condition is acyclic, see the following argument, developed after correspondence with Wlodek Rabinowicz. On the Minimal Structure Condition, for any populations X and Y where all individuals’ wellbeing levels are in the neutral range, a population X (with a number n of people in the neutral range) is more choice-worthy than a population Y iff X contains greater total welfare than the total welfare of any n people in Y.

Suppose, for reductio, that the Minimal Structure Condition were not cyclic, that is, on that Condition, there is a population P1 is more choice-worthy than a population P2, and P2 is more choice-worthy than a population P3, but P3 is more choice-worthy than P1. This means, by definition, that there is a population P1 (with $n1$ people) that contains greater total welfare than the total welfare of any $n1$ people in P2, but less total welfare than the total welfare of some $n1$ people in P3. P2 (with $n2$ people) contains less total welfare than the total welfare of some $n2$ people in P1, but more total welfare than the total welfare of any $n2$ people in P3. And, P3 (with $n3$ people) contains less total welfare than the total welfare of some $n3$ people in P2, but more total welfare than the total welfare of any $n3$ people in P1.

P3 can be the biggest population of this set, the smallest, or have same size as P1 or P2. Suppose P3 is the population with the biggest size. This means that the total welfare of the $n3$ people in P3 is smaller than the total amount of the $n2$ people in P2. But the total welfare of the $n2$ people in

P2 is smaller than the total welfare of any n_2 people in P1. Thus, the total amount of welfare of the n_3 people in P3 is smaller than the total amount of welfare of some n_2 people in P. Thus, P3 is less choice-worthy than P1: this contradicts the premise according to which P3 is more choice-worthy than P1.

Suppose P3 is the population with the smallest size. This means that the total welfare of the n_1 people in P1 is smaller than the total amount of the n_3 people in P3. But the total welfare of the n_3 people in P3 is smaller than the total welfare of any n_2 people in P2. Thus, the total amount of welfare of the n_1 people in P1 is smaller than the total amount of welfare of some n_2 people in P. Thus, P1 is less choice-worthy than P2: this contradicts the premise according to which P1 is more choice-worthy than P2.

Suppose P3 had the same size n_1 of P1. This means that the total welfare of the n_1 people in P1 is smaller than the total amount of the n_1 people in P3. The total amount of welfare of the n_1 people in P3 is smaller than the amount of welfare of some n_1 people in P2. However, any amount of welfare of any n_1 people in P2 is smaller than the total welfare of the n_1 people in P1. Thus, the total amount of welfare of the n_1 people in P3 is smaller than the total welfare of the n_1 people in P1: this contradicts the second sentence of this paragraph.

Suppose P3 had the same size n_2 of P2. This means that the total welfare of the n_2 people in P3 is smaller than the total amount of the n_2 people in P2. The total amount of welfare of the n_2 people in P2 is smaller than any amount of welfare of any n_2 people in P1. Thus, the total welfare of the n_2 people in P3 is smaller than some amount of welfare of some n_2 people in P1: this contradicts the second sentence of this paragraph.

Since all possibilities lead to a contradiction, the *reductio ad absurdum* is complete. I conclude that the Structured Range View is acyclic, and thus not susceptible of making incorrect judgements in cases of impartial information, such as the *Fickle Kidnapper* case.

Nor the Structured Range View is susceptible to the *Fickler Kidnapper* case, where the transitivity of “not worse than” fails. In fact, the *Fickler Kidnapper* case is not problematic for theories

where incommensurability is restricted to a range, and the Structured Range View limits incommensurable populations to the neutral range.

We can, however, worry that the Minimal Structure Condition implies a violation of transitivity *across* choice-sets. This can lead to the so-called “money pumps”, that I have mentioned in Chapter 4 and footnote 66. For example, some sequential choice problems, such as the ones I pointed out in 3.2, are a violation of transitivity of “at least as choice worthy as” across choice-sets. A simpler case of violation of transitivity across choice-sets is the following:

<i>Choice-set 1</i>	<i>Choice-set 2</i>	<i>Choice-set3</i>
A: 1 person at Ø6	B: no one	C: 1 person at Ø4
B: no one	C: 1 person at Ø4	A: 1 person at Ø6

In Choice-set 1, A is at least as choice worthy as B. In Choice-set 2, B is at least as choice worthy as C. On transitivity of “more choice worthy than” across choice-sets, C should be at least as choice worthy as A. But on the Structured Range View we must pick A over C: this contradicts transitivity of “better than” across choice-sets (in virtue of contradicting the transitivity of “more choice worthy than” in a context where goodness is the only variable).

However, I believe that any decision procedure as good as the Resolute Structure, mentioned in Section 3.2, should manage to bring transitivity also across choice-sets. Recall

Resolute Structure (in a choice sequence): When deciding among permissible populations consisting exclusively of neutral lives, once we choose a population P_1 of any size n at any choice-point c of a sequence of choices, we can, at any choice-point later than c of a sequence of choices, switch for another population P_2 only if P_2 features a set of lives (of size no greater than n) that contains at least as much total welfare as P_1 .

On Resolute Structure, every time we choose a population P_1 consisting of n people, we can switch it for another population P_2 only if P_2 has at least as much total wellbeing for a subset of people n as P_1 . Since the relation “at least as much total wellbeing as” is transitive, Resolute Structure ensures a transitive ranking when making a sequence of choices. So long as we are choosing among choice-sets in sequence, this ensures transitivity across choice-sets. It is hard to imagine a non-sequential choice across choice-sets.⁷¹ Thus, there is at least one decision procedure, namely Resolute Structure, where the Structured Range View is transitive both within and across choice-sets.

3.5 Out of the dilemma: Repugnance, Greediness, and the Structured Range View

It is relatively straightforward to show why my account does not imply what Thornley calls

Maximal Repugnance: Let life x and life y be lives that are identical, except that y has one fewer gumdrop’s worth of pleasure and one more hangnail’s worth of pain than x , such that (1) each population of wonderful lives is worse than some population of x lives and (2) each population of awful lives is better than some population of y lives. (Adapted from (Thornley, 2022, p.11))

In order to get Maximal Repugnance, the negative and the positive level need to be very close, and all critical range theories will avoid it. In particular, the Structured Range View avoids it because a gumdrop of pleasure and the hangnail pain are insufficient to fill the gap between \emptyset_0 and \emptyset_n (see Section 2.2).

⁷¹ One example would be voting theory, an area riddled by paradoxes independently from the Structured Range View: for examples and discussions of these paradoxes, see (Arrow, 1950; Gehrlein, 2006; Ottonelli, 2010; Pettit, 2001).

However, we want to avoid any kind of Repugnance. We want to ensure that the range between the zero level and the critical level is sufficiently wide to exclude Drab and Barely Conscious Lives.

In order to do so, the Structured Range View should assume the range I argued for in the last Chapter. In that Chapter I showed that, on the most prominent theories of when a life is good, bad or neutral, the neutral level spans from the Drab Life to the life in a coma, and indeed above the Drab Life (remember life in country B in the previous Chapter). Indeed it may reach even below the coma-life (see footnote 53). Thus, any instance of Repugnance is avoided by my theory.

Let us finally turn to the problem of Greediness. There are two versions of this problem. The most serious is:

Maximal Greediness: for any population of awful lives and any population of wonderful lives, (1) there is some population of straightforwardly-better-than-blank lives such that the population of awful lives is not worse than the population of wonderful lives plus the straightforwardly-better-than-blank lives, or (2) there is some population of straightforwardly-worse-than-blank lives such that the population of wonderful lives is not better than the population of awful lives plus the straightforwardly-worse-than-blank lives. (Thorley 2022, p. 17).

The Structured Range View avoids this problem. In this View, lives that are neutral only act as a tie-breaker for when there is perfect balance between good (or wonderful) lives and bad (or awful) lives.

To see why, let blank lives have some wellbeing level \emptyset_w between \emptyset_0 and \emptyset_n , let awesome lives have some wellbeing level w_a greater than \emptyset_n , and let awful lives have wellbeing level w_a smaller than \emptyset_0 , and let straightforwardly-better-than-blank lives have welfare w_b .

I will now show how my model avoids the disjunction (1) of Maximal Greediness. Consider any population P of any number x of people with welfare w_a , and another population Q of any number

y of people with welfare w_a , plus any number z of people with welfare w_b . The value of the populations, outside the range, is as follows: P has value $x \times w_{-a}$, which is a negative number (since x is positive and w_{-a} is negative), while Q has value $y \times w_a$, which is a positive number (since both y and w_a are positive). Any negative number is smaller than any positive number thus, on the Structured Range View, P is worse than Q. The z lives at w_b play no role, as lives outside the range already determine the betterness ranking.

Since P is any population of awful lives, and Q is any population of awesome lives joined with any population of straightforwardly-better-than-blank lives, disjunction (1) of Maximal Greediness is false. That is, it is false that, for any population of awful lives and any population of wonderful lives, there is some population of straightforwardly-better-than-blank lives such that the population of awful lives is not worse than the population of wonderful lives plus the straightforwardly- better-than-blank lives. The demonstration that my model avoids the disjunction (2) of Maximal Greediness is symmetrical. The Structured Range View avoids Maximal Greediness.

It is, however, interesting to notice that the model avoids Maximal Greediness even if we endorse a Broad Intuition of Neutrality Theory, and do not think that the range is bounded from above – that is, even if there is no upper limit \emptyset_n .

In fact, consider my previous example, where I had a population P of any number x of people with welfare w_{-a} , another population Q of any number y of people with welfare w_a , plus any number z of people with welfare w_b . If there is an \emptyset_0 , but no upper bound \emptyset_n of the neutral range, the (negative) value of a population is the total score below \emptyset_0 , and then we use what is above \emptyset_0 as a tie-breaker. P has a negative total score of $x \times w_{-a}$, Q has a negative total score of 0. Thus, P is worse than Q. Thus, if a Broad Intuition of Neutrality Theory endorses my model, it avoids the first conjunct of Maximal Greediness. The demonstration for the second conjunct is identical to the demonstration that the Structured Range View respects Benign Addition within the range, so I will not repeat it. We

conclude that a Broad Intuition of Neutrality Theory avoids Maximal Greediness by endorsing the model for the Intuition of Neutrality of the Structured Range View.⁷²

There is another version of the Greediness problem, the one originally formulated by Broome. As we have seen in Chapter 5, this version is more controversial (Rabinowicz 2009a). According to Broome, the following is a desideratum for any theory of population axiology:

Non-Greediness: “The net effect of one bad thing and one neutral thing should be bad”
(Broome 2004, p. 170)

On one interpretation, the Structured Range View is susceptible to Broome’s Greediness problem. In fact, it straightforwardly gets what Broome considers the wrong result in one of his cases mentioned in Section 3 of Chapter 5: Totalism: Between Repugnance and Greediness. Note that in that Section I mentioned two such cases: the Structured Range View is not susceptible to the Greediness case concerning climate change, as the Sample Quantity Condition rules it out.

I report Broome’s problematic case below, with the required modifications to notation. Consider four populations, with wellbeing distributions as follows:

Population A: $\emptyset 4, \emptyset 4, \emptyset 4, \dots \emptyset 4, \emptyset 6$, no one

Population B: $\emptyset 4, \emptyset 4, \emptyset 4, \dots \emptyset 4, \emptyset 6, \emptyset 1$

Population C: $\emptyset 4, \emptyset 4, \emptyset 4, \dots \emptyset 4, \emptyset 4, \emptyset 4$

Population D: $\emptyset 4, \emptyset 4, \emptyset 4, \dots \emptyset 4, \emptyset 4$, no one (Broome 2004, pg. 148/170)

Broome considers “Greedy” a theory allowing A to be as good as C. This is because:

⁷² The Broadest Intuition of Neutrality does not avoid Maximal Greediness, no matter which model it adopts.

Moving from A to C involves two things. First, the second-last person's wellbeing is reduced from $[\emptyset]6$ to $[\emptyset]4$. This is a bad thing. Second, an extra person is added at level $[\emptyset]4$. This is a neutral thing. The net effect of one bad thing and one neutral thing should be bad. But according to our theory, it is not bad; it is neutral. [Any theory with this implication does not represent] neutrality as it intuitively should be. It is a sort of greedy neutrality, which is capable of swallowing up badness or goodness and neutralizing it. (Broome 2004, 170).

The Structured Range View does imply that, in the choice-set of populations above, A and C (and *only* A and C) are permissible: in the choice-set there is no population (or subset of population) with as many people as A and greater total wellbeing, and the same is true for C. Thus, due to the Minimal Structure Condition, they are both permissible. Thus, on the Structured Range View, choosing between either A or C is equally rationally justified: none is better than the other.

However, the Structured Ranged Theory describes the situation a bit differently. It does not say that there is a change for the worse (because a person wellbeing's decreases) and one that is neutral (because a further person is added). Remember that, on Minimal Structure Condition, we can add m people to a population consisting of n neutral lives only if there is no population with $n+m$ people with greater total wellbeing than the resulting population. This means that one has to maximise total welfare relative to the number of people in the resulting population, even if this comes at the cost of lowering individual wellbeing to the n people.

This means what follows. On the Structured Range View, adding people with neutral welfare to a population is never required. However, it is permissible add people living neutral lives in a population only if this results in maximising the wellbeing of the resulting population. There are no requiring reasons to add people in the range, but there are justifying reasons to do so iff this comes with the maximisation of the total wellbeing of the resulting population. I expect Broome to agree that decreasing an individual's wellbeing is justified if this is necessary to increase a population's total wellbeing.

Thus, if we were in population A, it would not be required to increase the number of people in order to reach population C. However, it would be justified to do so. In fact, it is justified to decrease an individual's wellbeing if this maximises the total wellbeing of the resulting population C. Thus, on another interpretation, I think it is inappropriate to say that we have a change for the worse (the decrease in individual welfare) "swallowed" by a neutral change (the addition of a new person). The decrease in individual welfare is justified by the maximization of total wellbeing of the new population. It is still appropriate to interpret the decrease in individual welfare as a change for the worse, and, as prescribed by the Intuition of Neutrality, the creation of a neutral individual as a neutral change. However, there is another change: we cannot overlook that we obtain the greatest total wellbeing of the resulting population by decreasing individual welfare: this is a good thing, and is what justifies the bad thing (that is, it justifies the individual loss). So, a neutral thing *and a good thing* are justifying a bad thing. However, since the resulting population is neither better nor worse than the previous, this change is not required.

I think this second interpretation is much stronger than the first interpretation. The Structured Range View interprets as "change for the worse" only negative populations and shortfalls. There are no negative populations in the four populations of Broome's example. The shortfalls of the Structured Range are captured by violations of the Minimal Structure Condition. There are no such violations in the four populations. This is because, while there is a loss in individual welfare, there is also the maximization of the total wellbeing of the resulting population: this makes the resulting population better, despite the resulting population is not worse than the previous population. Within the terms of the theory, there is no neutral change swallowing a change for the worse.

The aforementioned example can be generalized. On the Structured Range View, any permissible addition of a new person in the range requires a maximization of the total wellbeing of the people who exist in the new population. In order to maximise total wellbeing, we are justified in decreasing wellbeing of people who existed independently of the permissible addition. Thus, any permissible addition creates justifying reasons to decrease welfare in another person, while not being

a required action. Any decrease in individual welfare within the range (any change for the worse) has to be justified in the maximisation of total wellbeing of the resulting population (a non-required change for the better).

Outside the range, the Structured Range View behaves like Totalism, and thus, it is not Greedy. Within the range, the only occasion where the Structured Range View allows for a decrease in individual's wellbeing to happen is when, by committing the permitted act of bringing a people into existence, we maximize total wellbeing. Thus, a decrease in individual wellbeing (change for the worse) is allowed only to preserve at least as much total wellbeing, which certainly balances the change for the worse. I think that the Structured Range View avoids any kind of Greediness and Repugnance, thereby solving the dilemma examined in Chapter 5: Totalism: Between Repugnance and Greediness.

I am, however, unsure whether John Broome would agree as to whether the Structured Range View respects Non-Greediness. In that case, at least the Structured Range view avoids the worst case of Greediness, namely Maximal Greediness, which is the thorniest horn of that dilemma.

To conclude, the Structured Range View respects most adequacy conditions for a theory of Population Ethics. It violates Non-Anti-Egalitarianism and related conditions, but it has a strong justification as to why it does so. We should accept the Structured Range View as a population axiology.

Let us now see how the theory fares when applied to aggregate value within lives. We will discover that it is equally powerful. As a bonus, the Structured Range View can explain why "Rollercoaster Lives" and "Very Short Lives" are better than "Drab Lives" or "Insect Lives".

4 Aggregating lives

4.1 The Structured Range View and Superiority views

Impossibility theorems in population axiology have almost perfect analogue theorems in aggregation of welfare within lives (Arrhenius, 2005; Cowen, 1996; Parfit, 2016; Pettigrew, 2019, p. 92; Rachels, 2004). Population axiologists have to deal with the Repugnant Conclusion, where for any population there is a better population consisting of sufficiently many people whose life is barely good. The analogous challenge for intrapersonal aggregation is to deal with the idea that, for any life, there is a better life that lasts sufficiently long, but each moment is just barely good. In Parfit's words,

Suppose that I can choose between two futures. I could live for another 100 years, all of an extremely high quality. Call this the Century of Ecstasy. I could instead live for ever, with a life that would always be barely worth living. Though there would be nothing bad in this life, the only good things would be muzak and potatoes. Call this the Drab Eternity.

On one view about what makes our lives go best [...] the Century of Ecstasy would have great value for me, this value would be finite, or have an upper limit. In contrast, since each day in the Drab Eternity would have the same small value for me, there would be no limit to the total value for me of this second life. This value must, in the end, be greater than the limited value of the Century of Ecstasy. (Parfit, 2004, pp. 160–161)

The tradition studying intrapersonal aggregation is much older than the tradition of population axiology. Famously, John S. Mill tried to defend that a pig's life, no matter how long, can never be as good as Socrates's life (Mill, 2000, p. 11). Identical cases can be found, for example, in (Plato, 1975, p. 21 b-d) and (McTaggart & Broad, 1968) among the most influential.

I defended a theory of population ethics that helps answering the impossibility theorems. In this Section, we see how the theory fares if adapted to aggregation of welfare within a life. We want a theory of aggregation of welfare within a life to avoid the implications that Centuries of Ecstasy are worse than the Drab Eternity, and that Rollercoaster Lives and Very Short Lives are better than Drab

Lives and Barely Conscious Lives. In this Section I show that slightly adapting my theory does the trick.

Recall that, in Chapter 6, I suggested it is not better or worse to continue one's life by moving to any of these countries:

Country A. You will live a Drab Life, with only muzak and potatoes, for the next six years.

Country B. You will live a Drab Life, with only muzak and potatoes, for the next *three* years. However, the muzak is just barely noticeably better recorded than in Country A, and the potatoes have some traces of salt that make them just noticeably tastier than in Country A.

Country C. You will live a Drab Life, with only muzak and potatoes, for the next *twelve* years. However, all muzak records stumble once each time they are played, and the potatoes are just noticeably staler than in Country A.

I suggested that, for any day in any of these countries, we may be indifferent as to whether to live this day rather than to skip it. It is important now to understand *why* it is not better nor worse to live any of these lives, why one may be indifferent between living a day in countries A to C or “skipping” it.

The most plausible answer seems to me that that all the welfare components in lives A to C are neutral themselves, despite some being better than others. We want slightly saltier potatoes rather than staler potatoes, but, at least under Parfit's description, we may be indifferent as to whether or not to skip them for their intrinsic properties (we still may want to eat them in order not to feel hungry). We want better recorded muzak rather than stumbling muzak, but we may be indifferent as to whether or not to skip our next muzak listening session.

The neutrality of these goods becomes more evident if we consider whether the *goods* involved in these examples have Positive or Negative Tradeability. They surely have not: if it is

rationally permissible to choose B or C, then an increase in quantity of potatoes does not compensate a decrease in quality, and an increase in quality of muzak does not compensate a decrease in duration, but they all remain rationally permissible choices.

We conclude that, if a life consists of only neutral welfare components, then that life is in the neutral range. As discussed in Section 3.3 of this Chapter, this suggests a sort of a “superiority” view (See Chapter 3), but with a crucial difference: Superiority views think that some positive goods are Superior to some others, while what I defend is that all positive goods are Superior to neutral goods. While the Superiority claim is puzzling among positive goods, my claim that all positive goods are Superior to neutral goods is so plausible that we may consider it an analytical truth, with no need of further justification. This makes the Structured Range View a particularly plausible candidate for a Superiority theory, if it is applied to both lives and populations, as I argue it should be in this thesis.

Let us now see how to order lives of different duration and momentary wellbeing according to the Structured Range View.

4.2 The Structured Range View applied to lives

In Chapter 6: Barely good lives and the Intuition of Neutrality we have established that there is a neutral range, where wellbeing levels are not good or bad for anyone. In Section 2 of this Chapter we called the upper limit of this range \emptyset_n , which is where Positive Tradeability stops applying, and the lower limit \emptyset_0 , which is where Negative Tradeability stops applying. The value of a life is the integral of wellbeing levels above \emptyset_n minus the integral of wellbeing levels below \emptyset_0 throughout a lifetime. The moments in the neutral range act as a tiebreaker, exactly as for populations.

When using the Structured Range View to choose between neutral populations, we needed to respect the Minimal Structure Condition, according to which it is permissible to pick a population of size n from a choice-set of neutral populations within the range only iff no other available population features a set of lives (of size no greater than n) that contains greater total welfare.

Likewise, when choosing among lives, or life continuations, within the neutral level, we need to respect

The Minimal Structure Condition (applied to lives): When choosing between lives (or life continuations) with all lifetime moments in the neutral range, choosing a life or life continuation (of length n) is permissible iff no other available life or life continuation features a set of lifetime moments in the neutral range (of length no greater than n) that contains greater total welfare.

Thus, in order to rank lives, we should adopt:

The Structured Range View (applied to lives): a life is better than another if it has greater total wellbeing outside the neutral range. When choosing among lives differing only with respect to lifetime moments within the neutral range, we use the Minimal Structure Condition. On this Condition, when choosing between lives (or life continuations) with all lifetime moments in the neutral range, choosing a life or life continuation (of length n) is permissible iff no other available life or life continuation features a set of lifetime moments in the neutral range (of length no greater than n) that contains greater total welfare.

And, in its axiological version,

The Axiological Structured Range View (applied to lives): a life is better than another if it has greater total wellbeing outside the neutral range. Within the range, a life (or life continuation) L_1 (of length n) is better than a life (or life continuation) L_2 iff L_1 contains greater total welfare than the total welfare of any amount of time n lived in L_2 . Two lives are equally good

if they contain the same amount of total welfare and are equally long. If two lives are neither better nor worse than one another, nor equally good, then they are incommensurable.

The evaluation of the Structured Range View for lives is identical to the evaluation of the Structured Range View for populations, and has many attractive features.⁷³ As the Structured Range View for populations does not imply the Repugnant Conclusion, the Structured Range View for lives does not imply that the Century of Ecstasy are better than Drab Eternity. Every second of the Centuries of Ecstasy is above \emptyset_n , while every second of the Centuries of Ecstasy is between \emptyset_n and \emptyset_0 .

If it is true that aggregating values within a life has exactly the same structural puzzles as aggregating value within populations, as assumed in (Arrhenius, 2005; Cowen, 1996; Parfit, 2016; Pettigrew, 2019, p. 92; Rachels, 2004), then my Structured Range View respects all desiderata for a theory of aggregation both within a life and within a population. The only condition I reject when aggregating within a life is the analogue for goods of Non-Anti-Egalitarianism, according to which, for any two lives L_1 and L_2 , if L_2 has higher average momentary wellbeing, higher total momentary wellbeing, more equally distributed wellbeing than L_1 , then L_2 is better.

This account also helps explaining why Rollercoaster Lives and Short Lived Lives are better than Barely Conscious Lives and Drab Lives. This is because Barely Conscious Lives and Drab Lives have no wellbeing level above \emptyset_n : it is permissible to get or refuse all the goods they consist in. On the other hand, Rollercoaster Lives and Short Lived Lives consist in goods that go beyond \emptyset_n and \emptyset_0 , thus it is better to have them than not to have them.

I conclude that the Structured Range View can avoid the intrapersonal analogue of the Repugnant Conclusion at the level of a life, and explain why Very Short Lives and (at least some) Rollercoaster Lives are better than Drab Lives or Insect Lives. Thus, the Structured Range View has

⁷³ The Structured Range View does not take into account non-aggregative variables in which a life can be better than another, such as for example the “shape” of a life (J. L. D. Brown, 2019; Dorsey, 2015; Velleman, 1991). These variables can be added on top of the structured range view, if one considers them relevant.

to be considered a promising account for how to aggregate lifetime wellbeing, and a particularly promising Superiority theory.

5. An alternative respecting Non-Anti-Egalitarianism

In this Chapter, I have defended what I take to be the best theory that violates Non-Anti-Egalitarianism. This theory is the Structured Range View, which is based on the Minimal Structure Condition. However, the Minimal Structure Condition by itself does not violate Non-Anti-Egalitarianism. Indeed, it can be a useful tool for all axiological theories, even the ones that are deeply committed to respecting Non-Anti-Egalitarianism. In this Section, I want to give an idea of the potential of the Minimal Structure Condition not applied to the Structured Range View.

In particular, the Minimal Structure Condition can be used to improve the standard “Critical Range” theories, illustrated in Chapter 6.⁷⁴ These theories respect Non-Anti-Egalitarianism, at least outside the Critical Range and between populations of the same size, but present some problems. For example, they violate Benign Addition, in the particularly vicious form of Maximal Greediness. Furthermore, they seem to provide insufficient guidance when choosing among some populations, thereby incurring in the (at least alleged) problem of Greediness. Broome suggests adopting a standard Critical Range View, and using the Minimal Structure Condition to solve these problems.

Remember that, on standard Critical Range Theories, there are many equally plausible zero levels. Thus, to rank any set of populations, there may be many equally plausible preliminary rankings, that differ as to which welfare level is the zero level. On standard Critical Range Theories, a population is better (worse, equally good) than another if it has greater (lower, equally as much) total wellbeing for each candidate zero level (i.e. each sharpening of the zero level). If a population is better for some rankings but not some other, then the two populations are incommensurable (see Chapter 5 for more on Critical Range Theories).

⁷⁴ Thanks to John Broome for this suggestion, in personal communication.

This theory respects Non-Anti-Egalitarianism. On Non-Anti-Egalitarianism, for any two populations P_1 and P_2 , if P_2 has higher average wellbeing than P_1 , higher total wellbeing than P_2 , and perfect equality, then P_2 is better than P_1 . On standard Critical Range Views, a population is better than another if it has higher total wellbeing for any interpretation of the neutral level, so it respects Non-Anti-Egalitarianism all cases where there is no disagreement among the rankings.

In all other cases, when there is disagreement among the rankings, it is the Minimal Structure Condition that decides which population is permissible and which is not. More precisely, when two populations are incommensurable, it is permissible to pick a population iff no other available population continuation features a set of people that contains greater total welfare. Of course, this theory needs to justify a “zero level” in order to calculate the total. I do not explore here which candidate would be best for such a theory: for the sake of illustration, I will assume here that we are considering the total is the sum of individual wellbeing if we consider the “zero level” as the lowest sharpening of the neutral level.

The Minimal Structure Condition *ensures further respect* of Non-Anti-Egalitarianism on this version of Critical Range View. Consider a population P_1 of size n_1 and total wellbeing tw_1 (where the total has as “zero level” the lowest sharpening of “zero level”), and a population P_2 with size greater than n_1 and greater total wellbeing than tw_1 . The presence of P_2 would make P_1 an impermissible to pick according to the Minimal Structure Condition,⁷⁵ even if P_1 and P_2 were incommensurate for the standard Critical Level View. By forcing to pick populations like P_2 over populations like P_1 , the Minimal Structure Condition makes Critical Level Views even more respectful of Non-Anti-Egalitarianism – indeed, a Critical Range View with the Minimal Structure Condition respects Non-Anti-Egalitarianism in all cases where one population has greater total and average welfare than another, and is perfectly equal, except in cases involving an empty population.

⁷⁵ Unless P_1 is the empty population: the empty population remains always permissible.

The Minimal Structure Condition improves Critical Range Views much further than that. Chiefly, it ensures they respect Benign Addition, and with it, it avoids Maximal Repugnance (but not Weak Maximal Repugnance). To see how standard Critical Range View violate Benign Addition, see Chapter 5, Section 4. To see how the Minimal Structure Condition respects Benign Addition, see Section 3.1 of this Chapter. Furthermore, if the Structured Range View respects Non-Greediness, as I argued for in Section 3.5, so may a Critical Range View improved by the Minimal Structure Condition, although the way justifying and requiring reasons work may require a different explanation, that I will not attempt here.

However, a standard Critical Range View improved by the Minimal Structure Condition has an important downside: it violates transitivity of “better than” (or rather, “more choice worthy than”) within the choice-set. To see why, consider what follows.

Suppose the Critical Range is between welfare levels 0 and 10. Suppose you can choose to create one between the following three populations.

A: 1bn people at 100

A+: 1bn people at 101, 100bn people at 1,

Z: 101bn people at 3

On standard Critical Range Theories, A is incommensurable with A+, A+ is worse than Z, and Z is incommensurable with A. The reader may remember why, but it is worth repeating.

A is incommensurable with A+ on standard Critical Range Theories, for the following reason. On the sharpening where wellbeing level 10 is the neutral level, A has value 90bn and A+ has value $(91\text{bn}-900\text{bn}) = -809\text{bn}$, so A is better than A+. On the sharpening where 0 is the neutral level, A has value 100bn and A+ has value 201bn, so A is worse than A+. Since the rankings disagree, A and A+ are incommensurable. It is an instance of the violation of Benign Addition, and an instance of Maximal Greediness, that standard Critical Range Theories get. However, in the version we are

examining in this subsection, the Minimal Structure Condition will rectify Benign Addition and dissipate Maximal Greediness.

A+ is worse than Z on standard Critical Range Theories, because standard Critical Range Theories respect Non-Anti-Egalitarianism in this instance. On the sharpening where wellbeing level 10 is the neutral level, A+ has value -809bn and Z has value -707bn. On the sharpening where wellbeing level 0 is the neutral level, A+ has value 201bn and Z has value 303bn. Thus, Z is better on any interpretation of the neutral level.

A is incommensurable with Z on standard Critical Range Theories, for the following reason. On the sharpening where wellbeing level 10 is the neutral level, A has value 90bn and Z has value -707bn. On the sharpening where 0 is the neutral level, A has value 100bn and Z has value 303bn. Since the rankings disagree, A and A+ are incommensurable.

These are all known implications of standard Critical Range Theories. Now, let us apply the Minimal Structure Condition to the incommensurable cases. On the Minimal Structure Condition, it becomes impermissible to pick A over A+: given the Minimal Permissibility Condition, A is impermissible as A+ has a subset of people of the same size as the A population, but with greater total welfare. We have respected Benign Addition and avoided Maximal Repugnance. However, when comparing A and Z, the Minimal Structure Condition says nothing: Z has greater total wellbeing than A, and A has a number n people with greater total wellbeing than any n people in Z.

Thus, A+ is more choice worthy than A, due to the Minimal Structure Condition. And, Z is more choice worthy than A+, given the standard Critical Range Theories ranking. If the theory respects transitivity, then Z is more choice worthy than A. However, neither the Minimal Structure Condition nor the standard Critical Range View gives us this result: both agree that Z and A are equally choice worthy. Thus, a Critical Range View improved by the Minimal Structure Condition violates transitivity of “better than”.

However, this theory is likely to be acyclic, and all the solutions I provided in section 3.4 should be available for the Critical Range View improved by the Minimal Structure Condition as

much as they are available to the Structured Range View. A more serious problem of The Critical Range View improved by the Minimal Structure Condition is Weak Maximal Repugnance. I consider the violation of Non-Anti-Egalitarianism around the neutral level a more acceptable cost to pay than Weak Maximal Repugnance.

In this subsection I showed that the Minimal Structure Condition has great potential, that is independent from the Structured Range View I defended in this Section. The Minimal Structure Condition can contribute positively to a population axiology even if we were to ultimately reject the Structured Range View. However, given the problems of this version of Critical Range View highlighted in this section, and until there is a solution to these problems, we should prefer the Structured Range View to this version of Critical Range View.

6. Conclusion

According to the Intuition of Neutrality, there is a range of wellbeing levels for lives that it is permissible to create or not create. This idea is thought to imply Greediness. In this Section I introduced a theory for population axiology, that I call the Structured Range View, according to which there is a range of lives such that adding any of them does not make a population better or worse, but it is impermissible to pick some sets of lives in the range over others. This respects the Intuition of Neutrality without implying a serious version of Greediness and respecting Benign Addition. However, the theory violates Non-Anti-Egalitarianism. This is intellectually costly, but not condemning, and the Structured Range View has a convincing justification for this violation, and indeed I suggest it may even be a desirable feature. The theory helps with aggregation within lives, too. We should consider the Structured Range View one of the most promising theories of population axiology available.

While the Structured Range View can respect most of the adequacy conditions thus far proposed in the literature, there are underexplored puzzles in population axiology to be solved. I discuss one of them in the next Section.

Chapter 8. The Monstrous Conclusion⁷⁶

Abstract

This paper introduces the Monstrous Conclusion, according to which, for any population, there is a better population consisting of just one individual (the Monster). The Monstrous Conclusion is deeply counterintuitive. I argue that the most plausible population axiology that does not imply the Monstrous Conclusion is Prioritarianism. The specific version of Prioritarianism I defend, which can avoid the Monstrous Conclusion, has not yet received any defense. According to this version of Prioritarianism, which I call Asymptotic Prioritarianism, there is diminishing marginal moral importance of individual welfare that can get close to, but never quite reach, some upper limit. I argue that Asymptotic Prioritarianism faces a theoretical cost, that I call the Absolute Priority Condition. However, the Absolute Priority Condition avoids the Monstrous Conclusion without appealing to implausible thresholds. Furthermore, I point out that the Absolute Priority Condition is an extreme version of what I call the Trade-off Condition, an already noteworthy problem facing other (more widely endorsed) versions of Prioritarianism. I conclude that it is better for a theory to imply the Absolute Priority Condition and avoid the Monstrous Conclusion than to imply the Monstrous Conclusion and the Trade-off Condition. The potential for Asymptotic Prioritarianism is substantial.

⁷⁶ A version of this chapter is forthcoming in Synthese.

1. Introduction

Consider the following:

The Monstrous Conclusion: for any population, there is a better population consisting of one individual.

The Monstrous Conclusion is deeply counterintuitive. I argue that we should concentrate our attention on theories that can avoid it.

The Monstrous Conclusion bears a striking resemblance to Nozick's Utility Monster objection (2012, p. 41), one of the most widely cited objections to utilitarianism (Briggs & Nolan, 2015; Kamm, 2015; Miller, 2021; Rosenqvist, 2020; Vallentyne, 1991). According to Nozick, 'Utilitarian theory is embarrassed by the possibility of utility monsters who get enormously greater gains in utility from any sacrifice of others than these others lose' (Nozick, 2012, p. 41). Nozick's case concerns a fixed population. The two possible populations are the status quo population and one in which everyone has their resources transferred to a 'Monster' and suffers the result. By contrast, the Monstrous Conclusion does not require the same people to exist in both populations: in one possible population there are, say, billions of very happy people, in another, there's just the Monster. While Nozick's Utility Monster is extremely influential, its variable-population analogue is underexplored.

The Utility Monster has been overshadowed by the Repugnant Conclusion in population ethics, in part because of how Parfit defended that the Repugnant Conclusion had independent interest despite Nozick's pre-existing Utility Monster. Parfit argues that Nozick's Utility Monster is possible only with major changes to the laws of nature, that it is "is a deep impossibility", and that we cannot successfully imagine it. Specifically, he asserts that the welfare of the Monster needs to be higher than what the laws of nature allow, and therefore "It seems a fair reply that we cannot imagine, even in the dimmest way, the life of this Utility Monster. And this casts doubt on the force of the example"

(1986, 389). However, it is not at all clear why the laws of nature should prevent someone from having an arbitrarily great amount of welfare.

Instead, according to Parfit, the Repugnant Conclusion is not a deep impossibility. According to the Repugnant Conclusion, for any population, there is some better population containing only lives that are barely worth living.⁷⁷ Parfit argues that “the difference [between the Repugnant Conclusion and the Utility Monster] is that the greater sum of happiness comes from a vast increase, not in the quality of one person's life, but in the number of lives lived. And [The Repugnant Conclusion] is neither deeply impossible, nor something that we cannot imagine. [...] So the example cannot be questioned as one that we can hardly understand.” (1986, 389)

By contrast, I contend that both the Repugnant Conclusion and the Monstrous Conclusion are similarly important for a satisfactory population axiology to avoid.

Specifically, I find Parfit's case for the Repugnant Conclusion being more important than Nozick's Utility Monster as a test for moral theories unconvincing. What I find unconvincing is not that the welfare of Nozick's Utility Monster may be hard to imagine. Rather, I find unconvincing that, as there are arguments for the unimaginability of Nozick's Utility Monsters, there are also arguments for the unimaginability of the Repugnant Conclusion of comparable strength. Indeed, one of the other rare uses of Nozick's Utility Monster in variable population context⁷⁸ has been to argue that we cannot imagine the enormous quantity of people necessary for the Repugnant Conclusion ((Ng, 1989, p. 242) ; for further discussion of the unimaginability of large numbers, see (Broome, 2004, p. 57; Gustafsson, 2022; Huemer, 2008, p. 904; Tännsjö, 2002)).

In response, many argue while we may not be able to grasp *how many* are people involved, we can grasp that *no number of people* would suffice (Mogensen, 2022; Parfit, 2016, p. 111; Pummer, 2013; Temkin, 2012, pp. 35, 121–122, 155). Similarly, while we may not be able to grasp *how good*

⁷⁷ This is a modification of the formulation in (Parfit 1984, p. 388). I assume here and throughout the paper that things other than welfare and number of people are equal across the populations I compare. For example, I assume that the two populations do not differ in desert, autonomy, genesis, aesthetic value, and so on.

⁷⁸ Others are (Pivato, 2014, 2018). They will be discussed in section 2.2.

the Monster's life is, we can still grasp that *there is no* amount of goodness for the Monster that would suffice. Moreover, even Parfit says that Nozick's Utility Monster 'may provide a partial test for our moral principles. We cannot simply ignore imagined cases' (1986, 389).

Meanwhile, others have argued for views where the life of the Utility Monster could also be rather mundane, for example on views where there is non-diminishing value to experiencing a given momentary quality of life for a longer time (Arrhenius, 2000, p. 44; McTaggart & Broad, 1968; Parfit, 2004, 2012). Moreover, even if the Repugnant Conclusion were compatible with the laws of nature but the Monstrous Conclusion were not, avoiding one but not the other would make our population axiology contingent on these facts in implausible ways (Arrhenius, 2000, pp. 50–51).

While enormous efforts have gone to exploring ways of avoiding the Repugnant Conclusion, there has been little discussion on how to avoid the Monstrous Conclusion (see (Blackorby et al., 2005; Ng, 1989; Parfit, 2016; Temkin, 2012; Thomas, 2018) and many others). Given how many ways there are that the Repugnant Conclusion and Monstrous Conclusion would be of comparable importance, there is clearly an imbalance between attempts to address the two. In this paper, I begin to tackle this imbalance.

I argue that a promising way to avoid the Monstrous Conclusion is Prioritarianism. Specifically, I show that the only kind of Prioritarianism which can avoid the Monstrous Conclusion is an until now undefended form of Prioritarianism according to which there is always moral importance to increasing an individual's welfare, but there is some limit, or asymptote, to the moral importance of one individual's welfare, while there is no asymptote to the moral importance of multiple other people's welfare. I call this Asymptotic Prioritarianism.

Nevertheless, every population axiology has its counterintuitive consequences, Asymptotic Prioritarianism included. Therefore, to establish what counterintuitive consequences the 'least-bad' population axiology must compromise on, more must be done to develop axiologies that can avoid the Monstrous Conclusion.

2. The Monstrous Conclusion

2.1 *The Monster*

According to the Monstrous Conclusion, for any population, there is a better population consisting of just one individual who is sufficiently better off (the Monster). The Monstrous Conclusion is deeply counterintuitive.

To appreciate the implausibility of this claim, consider that the Monstrous Conclusion applies to *any* population. Suppose we have a vast population, say a hundred billion people, each leading an excellent life. The lives could be as good as you like, much better than anyone has today. They enjoy great pleasures, have deep understanding, fulfil challenging projects, and develop meaningful relationships. They also do not experience pain or agony, there is no evil in the world and no-one is subject to malice or other's domination. Nevertheless, the Monstrous Conclusion implies that there is a *better* population consisting of a single individual, the Monster. I expect most people will find the Monstrous Conclusion deeply implausible, even impossible to believe.

While the Monstrous Conclusion has similarities with Nozick's Utility Monster objection, the Monstrous Conclusion enables us to more clearly identify the problems with giving the Monster too much moral importance (2012, p. 41). The key difference between the Monstrous Conclusion and Nozick's Utility Monster objection is that Nozick considers a fixed population case, whereas the Monstrous Conclusion is about variable populations. This difference has a number of important upshots.

Firstly, a number of negative features that one can appeal to in order to explain why all the resources should not be given to the Utility Monster in Nozick's objection are not available to resist the Monstrous Conclusion. For example, the rest of the population is harmed in Nozick's objection but not in the Monstrous Conclusion (Broome, 2004; Bykvist, 2007; Heyd, 1988; Parfit, 1986, pp. 487–489) - the Monster just exists by themselves, and no one needs to be harmed. Similarly, Nozick's case involves vast inequality between the Utility Monster and the rest of the population, whereas the Utility Monster exists alone in the Monstrous Conclusions case, so cannot involve any inequality

(Nebel, 2017, p. 898; Otsuka, 2012, p. 370; Voorhoeve & Fleurbaey, 2016). As the Monstrous Conclusion is highly implausible despite lacking these bad features, there must be sufficient explanation for the Monstrous Conclusion's implausibility that does not rely on these features.

Secondly, none of the standard positive features that people appeal to in order to rank populations straightforwardly count against the Monstrous Conclusion.⁷⁹ The population only containing the Monster has: higher total welfare (Arrhenius, 2000, pp. 37–51; Huemer, 2008; Parfit, 1986, pp. 397–389; Tännsjö, 2002); average welfare (Grill, 2023; Parfit, 1986, p. 387; Pressman, 2015); no inequality (Nebel, 2017, p. 898; Otsuka, 2012, p. 370; Voorhoeve & Fleurbaey, 2016); everyone who exists in both is much better off and no-one who exists in both is worse off (Heyd, 1988; Roberts, 2009; Ross, 2015; Temkin, 2012, sec. 12); there are more perfectionist goods (Beard, 2020; Parfit, 2016) and so on...

As the Monstrous Conclusion is deeply counter-intuitive, there must be more to the value of populations than the literature has assumed so far. Given that the Monstrous Conclusion is deeply counter-intuitive despite these differences, the Monstrous Conclusion is an excellent tool for understanding what is intrinsically unsatisfactory about the Utility Monster population.

2.2 Most theories imply the Monstrous Conclusion

Having motivated the independent interest in the Monstrous Conclusion, in this sub-section I explore how extant population axiologies fare with respect to avoiding the Monstrous Conclusion. After showing the limits of a seemingly simple solution for avoiding the Monstrous Conclusion, in this sub-section I show that the Monstrous Conclusion is implied by a range of major population axiologies. For reasons of space, I cannot go through all population axiologies defended in the literature. I therefore restrict my analysis to the Total View, the Average View, views combining Total and

⁷⁹ This will be explained in greater detail in section 2.2.

Average aspects, Critical Level and Critical Range Views, and Person Affecting Views. These are the views covered in a recent influential survey (Greaves 2017). I conclude this section by pointing out a view defended in the literature that does avoid the Monster, but does so in undesirable ways.

A seemingly simple, but ultimately unsatisfactory, way to escape the Monster is with an assumption of bounded utility. On the assumption of bounded utility, there is an upper limit to how much welfare any being can have. If there is a limit to how much welfare any being can have, then this also limits on how much the Monster can contribute to the value of a population – that is the contributory value of the monster to a population. According to the Monstrous Conclusion, for any population there is a better population consisting of only one individual with greater contributory value than the given population. If welfare is bounded above, and if, for some given population, the Monster needs to have a contributory value greater than the limit to be better than the given population, then the assumption of bounded utility prevents the Monstrous Conclusion.

Bounded utility is often used as a convenient assumption in economics. The primary justification given for bounded utility is that it is the easiest way to avoid certain ‘paradoxes’ that have unbounded payoffs, most famously the St Petersburg and Pasadena paradoxes, but adopting bounded utility is not the only solution to such paradoxes (Arrow, 1971, p. 92; Cowen & High, 1988; Nover & Hájek, 2004). See (Blackorby et al., 2005, p. 91; Kreps, 2013, pp. 11–13; Savage, 1972; Von Neumann & Morgenstern, 2007) for those who adopt bounded utility and (Buchak, 2017; Dillenberger & Vijay Krishna, 2014; Fishburn, 1976; Hájek & Smithson, 2012; Kosonen, 2022) who do not, among many others. As both bounded and unbounded utility are widely used in economics depending on the application of a given theory or model and population ethics is meant to be applied to any logically possible population (Arrhenius, 2000, pp. 31–35, Forthcoming, p. 41; Beckstead & Thomas, 2023; Gustafsson, 2020; Huemer, 2008; Thomas, 2018), population ethics should not merely assume bounded utility. Indeed, from the outset of the field of population ethics, unbounded utilities were accepted as at least a partial test of our theories (see Parfit, 1986, p. 389).

Potentially more significantly, some philosophical debates in prudential axiology have implications for whether welfare must have some upper limit. As an example of these debates, consider the debate between those who argued that increasing the duration of a happy life cannot improve this life beyond a certain limit (Beglin, 2017; Kagan, 2012; Smuts, 2011; Temkin, 2008; Williams, 1973, pp. 224–232) and those who have argued the opposite (Beckstead & Thomas, 2023, pp. 13–14; Bruckner, 2012; Fischer & Mitchell-Yellin, 2014; Gorman, 2017; Greene, 2017). I am personally unconvinced that any such arguments have established that individual welfare must have an upper limit. Nevertheless, I do not argue for that here, nor do I need to in order to assume unbounded utility.

Indeed, assuming unbounded utility has methodological advantages regardless of whether utility is actually bounded or unbounded. If utility is unbounded, then we have an account of population ethics that captures the full range of cases. If utility is bounded, then we still have a theory covering the full range of cases by considering the part of our theory up to that bound, but we can also determine which important features in population ethics depend on these controversial debates in prudential axiology. Therefore, whether I am right or not about whether utility is in fact bounded, significant results can be obtained by using an assumption of unbounded utility. Thus, I assume unbounded utility for the rest of this paper.

The Total View. According to the Total View, one population is better than another if the total amount of welfare is higher (Parfit 1984, 387-389; Arrhenius 2000, 37-51; Huemer 2008; Tännsjö 2004). The Total View implies the Monstrous Conclusion. For any population, that population will have a given amount of total welfare. So long as the Monster has more welfare than that, the population containing the Monster only is better according to the Total View. Therefore, for any population, there is a better population consisting of a single individual (the Monster): this is the Monstrous Conclusion. For more on the Total View, see Chapter 5: Totalism: Between Repugnance and Greediness.

The Average View. According to the Average View, one population is better than another if the average amount of welfare is higher (Parfit 1984, 387; Pressman 2015; Grill 2023). The Average View implies the Monstrous Conclusion. For any population, that population has some average welfare level. As the Monster is the only member of the single individual population, its welfare is the same as the average welfare of the population. So long as the Monster has more welfare than the average welfare of the given population, the population containing only the Monster is better according to the Average View. Therefore, for any population, there is a population consisting of a single individual that is better. This is the Monstrous Conclusion. For more on the Average View, see Section 3, Chapter 3. The Benign Addition Paradox .

Combined Views are a family of theories that aims to combine the Total View and the Average View, in such a way that the strengths of each view cover the other view's weaknesses. Typically, these theories tend towards the Total View when evaluating small populations, and towards the Average View when evaluating larger populations (Hurka, 2011, pp. 25–36; Ng, 1989; Sider, 1991). Since the Monstrous Conclusion is a weakness of both the Total View and the Average View, Combined Views inherit this problem: for any population, there is a Monstrous Population with greater total and average welfare.

Critical Level and *Critical Range Views* are modifications of the Total View aimed at preventing populations containing only barely worth living lives having unbounded value. These views either adopt a sufficiently good level, where lives that are not sufficiently good count against a population (these are Critical Level Views: see (Blackorby et al., 1997, 1998, 2005)) or an extended range around the neutral level that does not count in favour of a population (these are Critical Range Views: see (Gustafsson, 2020; Rabinowicz, 2009, 2022; Thornley, 2022)). For more on Critical Range Views, see Chapter 5. I defended a Critical Range View in Chapter 7.

However, as these views only adjust the contributory value of welfare around the neutral level, they do nothing to prevent the welfare of one individual having unbounded contributory value. Therefore, for any population, that population will have a given total contributory value adjusted

by the critical level or range. So long as the Monster has more contributory value than that (adjusted by the critical level or range), the population with only the Monster in it is better according to Critical Level and Critical Range Views. Thus, for any population, there is a better population consisting of a single individual (the Monster): this is the Monstrous Conclusion.

Person Affecting Views are a family of views according to which welfare that involves making people better and worse off has a distinctively central moral importance compared to making people who have positive welfare (Heyd, 1988; Roberts, 2009; Ross, 2015; Temkin, 2012, sec. 12).⁸⁰ *Person Affecting Views* imply the Monstrous Conclusion. As the Monstrous Conclusion just requires that there is *some* one-person population which is better, let us suppose that the individual is one of those who exists in the given population. Now suppose that that person gets much more welfare than the whole of the given population put together. As that person is much better off, there is enormous comparative benefit to producing the one individual population, as well as much greater welfare overall. Thus, as there is both greater comparative benefit and welfare overall, according to *Person Affecting Views* the one individual population is better: this is the Monstrous Conclusion.

To diagnose the implausibility of the Monstrous Conclusions we must look beyond these major population axiologies. I begin by considering and rejecting Sufficiencyarianism before offering my own Prioritarian theory that avoids the Monstrous Conclusion.

Sufficiencyarianism is the family of views which claim that absolute priority should be given to the welfare of people below a certain welfare threshold (Bossert et al., 2022; Brown, 2005; Casal, 2007; Crisp, 2003; Frankfurt, 1987; Hirose, 2016). Sufficiencyarianism captures the attractive idea that it is more important to make people sufficiently well off than to make the already well off even better off, albeit in an extreme form. This idea can be used to resist the Monstrous Conclusion by arguing that there is absolute priority to creating people who are sufficiently well off, rather than making the monster better and better off.

⁸⁰ See (Arrhenius 2000, 114-138; Greaves 2017, 11-16) for more information and some important distinctions on *Person Affecting views*.

For example, on the sufficientarian “head-count approach”, the aim is to ensure that “as many people as possible have enough” (Frankfurt 1987 p. 31). That is, populations are ranked according to the number of people who have enough, therefore, as the Monstrous population only has one person who has enough, any population with at least two people who have enough is better than the Monstrous population.

There is a sense in which avoiding the Monstrous Conclusion involves an ‘absolute’ priority claim – the welfare of sufficiently many sufficiently well off people has priority over the welfare of one individual, no matter how much welfare is at stake. Nevertheless, I shall argue that Sufficientarianism faces fatal defects that can be avoided by capturing this ‘absolute’ priority claim in the form of Prioritarianism instead.

In particular, Sufficientarian theories face a serious, well-known problem. These theories rely on a sufficiency threshold below which any welfare increase gets absolute priority over welfare increases above the threshold: this creates a particularly extreme kind of discontinuity in the welfare spectrum. Sufficientarianism implies that there are two welfare levels w_1 and w_2 , arbitrarily close to one another, but w_1 just below the threshold and w_2 just above, such that an arbitrarily small improvement to w_1 is more important than an arbitrarily great improvement to w_2 . Since w_1 and w_2 can be arbitrarily similar, it is hard to believe that we should treat them so dissimilarly.

As there is no plausible candidate for a point in the welfare spectrum that demarcates such a stark difference, sufficientarianism’s requirement of this discontinuity in the welfare spectrum is widely considered implausible (Arneson, 2000, p. 56, 2002, p. 194; Casal, 2007, p. 317; Dorsey, 2014, pp. 50–53; Holtug, 2010, pp. 207, 227–31; Timmer, 2022, pp. 308–309).

I argue that the best theory in population axiology that avoids the Monstrous Conclusion is a version of Prioritarianism. The specific version of Prioritarianism which can avoid the Monstrous Conclusion has not yet received any defense and is the focus of the next Section.

3 *Prioritarianism*

3.1 *Asymptotic Prioritarianism*

Prioritarianism's central claim is that there is greater moral importance, or priority, to giving welfare increases to the less well off. In other words, there is less moral importance to giving welfare increases to the better off (Parfit 1991; 2012; Holtug 2017; Adler 2021).

While Parfit intended Prioritarianism to be limited to fixed populations (Parfit, 2012, p. 440; Segall, 2022), prioritarianism, including the kind of Prioritarianism I am about to defend in this section, is compatible with many population axiologies. For now, let us assume its most popular version (Adler, 2019; Holtug, 2017, 2022; Segall, 2022), Total Prioritarianism, according to which population value is the total sum of individual priority-adjusted welfare (we will see other prioritarian population axiologies in the next section).

To better understand the arguments of this section, some formalization can be useful. Total Prioritarianism can be expressed as ranking outcomes according to the formula $\sum_i g(w_i)$, where w_i is the amount of welfare of each individual i , and $g(\cdot)$ is a concave transformation function. A concave transformation function is a function where the gradient of the slope is always decreasing. Given that the gradient corresponds to the contributory (or marginal) value of each increase in welfare, a decreasing gradient represents a decrease in value of additional welfare increase. In other words: the lower an individual's welfare, the more an increase to that individual's welfare contributes to population value.

Prioritarianism is closely related to the Pigou-Dalton principle, the foundation of economic work on inequality (Adler 2013). The Pigou-Dalton principle states that any non-rank-switching fixed transfer from the better off to the less well-off makes a population better. By "non-rank-switching" I mean that "the one who starts out with less does not end up with more than the other" (Adler 2013, pg 1). Any concave prioritarian transformation function will satisfy the Pigou-Dalton principle, as

the average gradient for the loss of welfare for the better off individual will be less than the average gradient for the gain of welfare for the less well-off individual.

In addition to fixed transfers, the foundational motivations for Prioritarianism also motivate to transfer welfare from the better off to the worse off even if a fraction of the transferred welfare is lost. This is because we may consider it sufficiently more important that the less well-off individual gets the welfare increase that some loss of total welfare can still result in a better population overall. We may believe that the priority of the worse off is not trivial with respect to wellbeing, it is not just a tie breaker.

For example, if we were to say that it is better to transfer 101 units of welfare from someone with 1,000 units to someone with 10 units, but that it would *not* be better to take 101 units from someone with 1,000 units to give someone starting at 10 an additional 100, then it seems that we would be Prioritarians in name only – there could only be at most trivial amounts of priority between the individuals. There must be some specific sufficiently small amount of the transfer in wellbeing from the better off to the worse off that we must be willing to pay to make the transfer.

This core prioritarian intuition applies so long as one of the individuals is sufficiently better off than the other. Not only it is better to take 101 from someone with 1000 units to give 100 units to someone with 10 units, but also better to take 101 from someone with 100,000 units to give 100 units to someone with 1000 units and so on. To properly respect the core prioritarian intuition, one must satisfy what I call “Non-trivial Priority throughout the Welfare Range”.

Non-trivial Priority throughout the Welfare Range says that there is at least some multiplier of welfare increases such that, for any welfare increase to a less well off individual, there is some better off individual such that it is better to give the initial welfare increase to the less well off individual than the multiplied welfare increase to the better off individual. For example, in the illustrative cases, there was a multiplier of 1.01 such that for a welfare increase (e.g. 100 units) for the less well off individual, it is better that the less well off individual receives the given welfare increase than someone who is 100 times as well off receives the multiplied welfare increase

(1.01*100=101 units). To reject *Non-trivial Priority throughout the Welfare Range* is to fundamentally violate the core Prioritarian intuition.

Now I state *Non-trivial Priority throughout the Welfare Range* more formally. Let w_j and w_i be two welfare levels, m a multiplier greater than 1, Δw a change in welfare, and $g(\cdot)$ the prioritarian concave transformation function.

Non-trivial Priority throughout the Welfare Range: $\exists m > 1$ s. t. $\forall \Delta w$ and $\forall w_i, \exists w_j > w_i$ s. t. $g(w_i + \Delta w) - g(w_i) > g(w_j + m\Delta w) - g(w_j)$

There is some multiplier m of the welfare increases, no matter how close to 1, such that for any size of welfare increase Δw , and any welfare level w_i , there is some welfare level w_j of a person who is (way) better off, where it is better if a less well off individual at welfare w_i receives the welfare increase Δw than if a better off individual at welfare level w_j receives the multiplied welfare increase $m\Delta w$.

While intuitive, *Non-trivial Priority throughout the Welfare Range* does come with an unavoidable cost, as multiple iterations of it imply what I call the Trade-off Condition.

Roughly speaking, where Non-trivial Priority requires that we should be willing to lose some welfare in order to transfer welfare between two sufficiently differently off individuals, the Trade-off Condition states that *any* amount of welfare will justifiably be lost in order to transfer welfare between two sufficiently differently off individuals.

More precisely, let Δw be welfare difference, let w_i and w_j be wellbeing levels, and let k be multiplier.

Trade-off Condition: $\forall k > 1 \forall \Delta w$ and $\forall w_i, \exists w_j > w_i$ s. t. $g(w_i + \Delta w) - g(w_i) > g(w_j + k\Delta w) - g(w_j)$

The Trade-off Condition means that, for any multiplier k , any welfare increase Δw and any welfare level w_i , there is some sufficiently better off welfare level w_j such that giving the welfare increase Δw to the person at the given welfare level w_i is better than giving the multiplied welfare increase to the better off individual (even though the multiplier may be arbitrarily large).

To see that *Non-trivial Priority throughout the Welfare Range* implies the Trade-off Condition, consider the following argument.

By *Non-trivial Priority throughout the Welfare Range*, there is some multiplier $m > 1$ such that, for any welfare increase Δw and any initial welfare level w_0 , there is some better off individual at some welfare level w_1 such that it is better to give the given welfare increase Δw to the person at the given welfare level w_0 than the multiplied welfare increase $m\Delta w$ to the better off individual at w_1 .

As *Non-trivial Priority throughout the Welfare Range* applies to any initial welfare level and size of welfare increase, we can apply the principle to welfare level w_1 and welfare increase $m\Delta w$. By *Non-trivial Priority throughout the Welfare Range*, there is some better off individual w_2 such that it is better to give the given welfare increase $m\Delta w$ to w_1 than the multiplied welfare increase $m^2\Delta w$ to w_2 .

More generally *Non-trivial Priority throughout the Welfare Range* implies that there is a sequence of welfare levels such that for n iterations, it is better to give $m^n\Delta w$ welfare increase to w_n than welfare increase of size $m^{n+1}\Delta w$ to someone at welfare level to someone w_{n+1} . By transitivity, it is better to given welfare increase of size Δw to someone at w_0 than to give someone a welfare increase of size $m^n\Delta w$ at welfare level w_n . However, as this is true for all n , and because $m > 1$, as n tends to infinity, m^n also tends to infinity. In other words, m^n is unbounded. Therefore, for every $k > 1$, there exists some n such that $m^n > k$. As more welfare is always better, for such an n , giving a welfare increase of size $m^n\Delta w$ is better than giving a welfare increase of size $k\Delta w$.

Finally, we can see that, for any size of $k > 1$, there is some step in the sequence such that, it is better to give a welfare increase of size $m^n \Delta w$ than of size $k \Delta w$ to w_n but better still to give welfare increase of size Δw to welfare level w_0 - an instance of the Trade-off Condition. As Δw and w_0 were chosen arbitrarily, the Trade-off Condition applies to all welfare increases and initial welfare level.

I find the Trade-off Condition troubling – that there is no bound to the amount of welfare that might be sacrificed for the sake of priority is not a feature that would attract me to a theory. Nevertheless, when required to choose between satisfying *Non-trivial Priority throughout the Welfare Range* and avoiding the Trade-off Condition, I accept the Trade-off Condition as a necessary cost of satisfying our core Prioritarian intuition.⁸¹

Having argued that, as Prioritarians, we ought to already be committed to the *Trade-off Condition* for reasons independent of the Monstrous Conclusion, I now argue that Asymptotic Prioritarianism's strength above and beyond the *Trade-off Condition* is a small price to pay to avoid the Monstrous Conclusion.

Asymptotic Prioritarianism is a form of Prioritarianism where the concave transformation function $g(\cdot)$ has an upper bound, or asymptote, corresponding to some population which is better than any one individual population. The contributive value of one individual can always increase the more their welfare increases, just not beyond some limit corresponding to some population which is better than any one individual population.

Consequently, as the Monster gets better and better off, their contributory value gets closer and closer to the asymptote. Therefore, as the contributory value that the Monster would have with any further welfare increase is bounded above by the asymptote, the moral importance of giving the monster any further welfare increase tends to 0, if compared to increases for less well off individuals. In other words, Asymptotic Prioritarianism implies the following Absolute Priority Condition. Let w_i and w_j be two welfare levels, Δw be an increase in welfare, and k a multiplier.

⁸¹ See (Nebel & Stefánsson, 2023) for a discussion of other 'calibration' problems that affect Prioritarian theories.

Absolute Priority Condition: $\forall w_i, \forall \Delta w$ and $\exists w_j > w_i$ s. t. $\forall k > 1, g(w_i + \Delta w) - g(w_i) > g(w_j + k\Delta w) - g(w_j)$

The *Absolute Priority Condition* says that, for any amount of moral importance generated by giving some welfare increase Δw to some individual at welfare level w_i , there is some vastly better off individual at welfare level w_j such that, no matter how big the multiplier k of the welfare increase Δw is, it is more important to give Δw to w_i , than $k\Delta w$ to w_j . In other words, there is a form of absolute priority where it is better to give some welfare increase Δw to a given individual than to give any size of welfare increase $k\Delta w$ to someone who is sufficiently better off (since the better off person is so close to the asymptote, no increase for them could be of more moral importance than the given welfare increase to the given individual).

As with the Trade-off Condition, I take the Absolute Priority Condition to be a counterintuitive implication of an axiology – because of the same central feature that there is no bound to the amount of welfare that might be sacrificed for the sake of priority. The key difference between the Trade-off Condition and the Absolute Priority Condition is just one of quantification order. The Trade-off Condition chooses the arbitrarily large amounts of welfare to be outweighed by priority first, and then ‘finds’ a sufficiently well off person where giving that amount of welfare to them is not better. By contrast, the Absolute Priority Condition chooses something of importance to a given individual first, and then ‘finds’ a sufficiently well-off person where there is no amount that could be given to the better-off person which is more morally important than the given thing of importance to the given person.

While both conditions imply that an arbitrarily large amount of welfare can be outweighed by priority, only the Absolute Priority Condition, which puts a bound to the importance of increasing welfare level, is sufficient to ensure that the population which is better than the population only containing the monster – that is, is sufficient to avoid the Monstrous Conclusion. In some ways,

without committing to any arbitrary threshold, it captures the sufficientarian intuition that, if we have a significant welfare increase that we can give to some less well-off individual, we do not need to know how much welfare is at stake for some sufficiently better off individual, there is a form of ‘absolute’ priority for this welfare increase to this less well-off individual. When working out whether to give a meal to a hungry child, do we really need to know how much better off Scrooge McDuck could be made with the same amount of resources?

We have seen that the Absolute Priority Condition is a more extreme version of the Trade-off Condition. However, if we are prioritarians, we already embrace the Trade-off Condition, and thus, are already committed to the idea that there is no bound to the amount of welfare that might be sacrificed for the sake of priority. As prioritarians, we have two options: either accepting only the less extreme Trade-off Condition together with the Monstrous Conclusion, or avoiding the Monstrous Conclusion and accepting the more extreme Absolute Priority Condition. It seems to me that accepting the Absolute Priority Condition rather than the less extreme Trade-off Condition is an absolutely acceptable price to pay to avoid the Monstrous Conclusion. To the extent in which we find plausible that there is no bound to the amount of welfare that might be sacrificed for the sake of priority, we should consider Asymptotic Prioritarianism a promising theory for avoiding the Monstrous Conclusion.

However, prioritarianism is considered vulnerable to a particularly implausible version of the Repugnant Conclusion, according to which, for any population A, there is a better Z population consisting of people whose lives are barely good, even if the total welfare of Z is smaller than the total welfare than A (Adler, 2019; Holtug, 2017, 2022; Segall, 2022). Thus, at this point, some may object that, by accepting Asymptotic Prioritarianism, we are avoiding the Monstrous Conclusion by implying a worse version of the Repugnant Conclusion. In the next section, I will show that this is not the case: Asymptotic Prioritarianism can be implemented on different aggregative theories, and some implementations of Asymptotic Prioritarianism can avoid both the Monstrous and the Repugnant Conclusion.

3.2 Different Asymptotic Prioritarianism axiologies

Prioritarianism is a view about the importance of individual wellbeing, and by itself it is not a full population axiology. Parfit considers Prioritarianism a theory for fixed populations, and suggests to ‘quarantine’ it for variable populations precisely because of its implications when combined with other theories in population axiology – notably the Total View (Parfit, 2012, p. 440; Segall, 2022). However, Prioritarianism, and Asymptotic Prioritarianism in particular, can be combined with a wide range of population axiologies. While some combinations are less promising, as Asymptotic Prioritarianism exacerbates existing objections to those population axiologies, other combinations allow for the benefits of both components of the resulting population axiology and, finally, in some cases, Asymptotic Prioritarianism can also help mitigate the existing objections to those views.

The most discussed Prioritarian axiology is Total Prioritarianism, where first each individual’s welfare is priority-adjusted to give each the moral importance, or more precisely contributory value, of each individual’s welfare, and then populations are ranked according to the total of these amounts of contributory value (Adler, 2019; Holtug, 2017, 2022; Segall, 2022). However, while Total Asymptotic Prioritarianism can fare better than other forms of Total Prioritarianism as it can avoid the Monstrous Conclusion, Total Asymptotic Prioritarianism shares the problems of other forms of Total Prioritarianisms in that it exacerbates an existing objection to the Total View.

To see that Total Asymptotic Prioritarianism can avoid the Monstrous Conclusion is relatively straightforward. First, Asymptotic Prioritarianism implies that there is some limit of contributory value that the Monster cannot exceed. Given that lives worth living have some positive individual contributory value, Total Asymptotic Prioritarianism implies that a sufficient number of those lives will have greater total contributory value than the limit the Monster cannot exceed and, therefore, have greater total contributory value than any Monster. Such a population would be better than any one individual population, thus avoiding the Monstrous Conclusion.

However, Total Asymptotic Prioritarianism has an even worse form of a classic objection to the Total View, namely the Repugnant Conclusion. The Repugnant Conclusion states that, for any population, there is some better population consisting only of lives that are barely worth living. Total Asymptotic Prioritarianism implies an especially bad form of the Repugnant Conclusion. For any population, that population has some amount of total contributory value. Barely worth living lives will have some lower amount of contributory value but sufficiently many of them will have greater total contributory value than any given level. The difference with a classical Total View is that, because of the greater contributory value of the welfare for the less well off than the better off, even fewer barely worth living lives are required to have greater total contributory value than is the case with the classical Total View.

Given its serious problems in dealing with the Repugnant Conclusion, Asymptotic Total Prioritarianism cannot be considered a promising theory for population ethics, even though it avoids the Monstrous Conclusion. Fortunately, Asymptotic Prioritarianism can be combined with population axiological theories that avoid do Repugnant Conclusion.

The most discussed population axiology that avoids the Repugnant Conclusion is the Average View. However, a combination of the Average View and Asymptotic Prioritarianism is not a promising candidate. While there might be some benefits to combining Asymptotic Prioritarianism with the Average View, such an Average Asymptotic Prioritarianism would be unable to avoid the Monstrous Conclusion, as the welfare of any given population is lower than the average welfare of the Monster alone.

A far more promising candidate is the combination of Asymptotic Prioritarianism with Critical Level and Critical Range Views, a set of views with considerable existing support level (Blackorby et al., 1997, 1998, 2005; Gustafsson, 2020; Rabinowicz, 2009a, 2022b; Thornley, 2022), that the Structured Ranged View belongs to. Critical Level and Critical Range Asymptotic Prioritarianism take into account both prioritarian considerations for the welfare of the less well off compared to the better off, as well as requiring that welfare levels must be above a critical level or range in order to

have a positive contribution to population value. Critical Level and Critical Range Asymptotic Prioritarianism rank populations according to their total welfare, adjusted both by critical level (or range) and asymptotic prioritarian factors.

Critical Level and Critical Range Asymptotic Prioritarianism (including the Structured Range View) can avoid the Monstrous Conclusion because there is some limit to the contributory value of one individual's welfare. Given that the welfare of individuals above the critical level or range has positive contributory value, sufficiently many of them will have greater total contributory value than the limit of the contributory value of one single individual's welfare. Hence, that population will be better than any Monster, according to Critical Level and Critical Range Asymptotic Prioritarianism.

Moreover, Critical Level and Critical Range Asymptotic Prioritarianism can avoid the Repugnant Conclusion. For any population with all individuals above the critical level or range, that population will have positive total contributory value. By contrast, any population consisting only of individuals with lives that are barely worth living will be below the critical level or range (or within the range) and therefore not have positive total contributory value. Hence, these populations are not worse than any population consisting only of individuals with lives that are barely worth living – thus avoiding the Repugnant Conclusion. As I argued in Chapter 5: Totalism: Between Repugnance and Greediness, however, the way classical Critical Range Views avoid the Repugnant Conclusion is unsatisfying, and Critical Level Views violate the Non-Sadism condition illustrated in Chapter 3. The Benign Addition Paradox : only the Structured Range View (and possibly the suggested modification to standard Critical Range Views in Chapter 7, Section 5) can satisfactorily avoid the Repugnant Conclusion.

Critical Level and Critical Range Asymptotic Prioritarianism is therefore able to combine the most important benefits of both Asymptotic Prioritarianism and Critical Level or Critical Range Views, and chiefly of the Structured Range View. This does not mean that we should necessarily endorse Structured Range Asymptotic Prioritarianism in population ethics, as we may find the Trade-

off Condition more problematic than the Monstrous Conclusion. Indeed, the Trade-off Condition seems to me an unjustified instance of hypersensitivity, and I am skeptical we should adopt it.

Another promising candidate is the combination of Asymptotic Prioritarianism with Combined Views. Combined Asymptotic Prioritarianism combines asymptotic diminishing contributory (or marginal) value for both additional welfare and additional individuals at a given welfare. Combined Asymptotic Prioritarianism is thus able to avoid both the Monstrous Conclusion and Repugnant Conclusion. To see this, consider the following.

No matter how many individuals are added with barely worth living lives, the contributory value of the population cannot exceed the limit for barely worth living lives. As the limit is greater for higher amounts of welfare, there is a population containing individuals with the higher amount of welfare which is better than all populations consisting of barely worth living lives. Similarly, no matter how much one individual's welfare is increased, the contributory value of the population cannot exceed the limit for one individual. As the limit for more individuals is higher, there is a population with more people which is better than any population consisting of only one individual. Therefore, Combined Asymptotic Prioritarianism avoids both the Monstrous Conclusion and Repugnant Conclusion.

Moreover, unifying Combined Views with Asymptotic Prioritarianism may mitigate the existing objections against Combined Views. For example, Sider himself is skeptical of his own Combined View since "it generates rather extreme results with respect to distributive justice" (Sider 1991, note 18). However, when combined with Asymptotic Prioritarianism, there would be a significant corrective to these distributional concerns.

To conclude, despite my personal skepticism towards prioritarianism in general, motivated by the Trade-off Condition, the potential of Asymptotic Prioritarianism in population ethics is enormous: further research is required in population ethics on Trade-off Conditions and Absolute Priority Conditions in Prioritarianism and how Asymptotic Prioritarianism can combine with other theories in population axiology to mitigate existing problems.

4. Conclusion

In this Chapter, I introduced the Monstrous Conclusion, a disturbing implication of most theories of population ethics – that for any population there is a better population consisting of one individual. The Monstrous Conclusion is roughly as disturbing as the Repugnant Conclusion, but has received far less attention, and examining it may clarify our intuitions concerning Nozick’s more famous Utility Monster. The Monstrous Conclusion is implied by most prominent theories of population ethics (total and average theories, theories combining total and average aspects, person affecting theories).

I argued that there is one hitherto undefended form of Prioritarianism that is uniquely well equipped to avoid the Monstrous Conclusion. This form of Prioritarianism is a member of what I call the family of Non-Trivial Priority theories, according to which it is better to transfer welfare from the better off to the worse off even if some portion of that welfare is lost. Particularly, I show that, on Non-Trivial Priority theories this lost portion of welfare has to get bigger as the two welfare levels are distant: this implies what I call the Trade-off Condition, for any amount of moral importance generated by giving some welfare increase Δw to some individual at welfare level w_i , there is some vastly better off individual at welfare level w_j such that, no matter how big the multiplier m of the welfare increase Δw is, it is more important to give Δw to w_i , than $m\Delta w$ to w_j . However, who takes seriously the priority view has to agree that the “Trade-off Condition” is a price we should be willing to pay.

The form of Non-Trivial Priority that best avoids the Monstrous Conclusion is what I call Asymptotic Prioritarianism, according to which the priority of the welfare of well off people over less well off people cannot exceed some asymptotic limit. This limits how much priority a single life can get in a population, but does not limit how much priority multiple people can get in a population: this enables Asymptotic Prioritarianism to avoid the Monstrous Conclusion.

Avoiding the Monstrous Conclusion, too, comes at an intellectual cost: I call it the Absolute Priority Condition, according to which for any multiplier k , any welfare increase Δw and any welfare level w_i , there is some sufficiently better off welfare level w_j such that giving the welfare increase Δw to the person at the given welfare level w_i is better than giving the multiplied welfare increase to the better off individual (even though the multiplier may be arbitrarily large). However, Asymptotic Prioritarianism pays this cost most acceptably, as the Absolute Priority Condition, necessary to avoid the Monstrous Conclusion, is just a more pronounced version of the Trade-off Condition already implied by Non-Trivial Priority theories. However, by itself Asymptotic Prioritarianism is not a complete population axiology, it must be combined with some further form of aggregation.

I show that combining Asymptotic Prioritarianism with other population axiologies can enable us to combine the positive features of each, although it may not mitigate or even exacerbate their other flaws. For example, although Total Asymptotic Prioritarianism can avoid the Monstrous Conclusion, it faces an even worse form of the Repugnant Conclusion. By contrast, a Structured Range Asymptotic Prioritarianism is able to convincingly avoid both the Monstrous Conclusion and the Repugnant Conclusion.

Crucially, I do not argue that we should endorse Asymptotic Prioritarianism as a result of the arguments in this paper. I have deep reservations about whether the Trade-off Conditions facing Non-Trivial Priority theories can be adequately addressed or explained away and existing population axiologies all have well-known counterintuitive consequences. However, avoiding the Monstrous Conclusion is a problem that population axiologies must take seriously and the potential of Asymptotic Prioritarianism, particularly Structured Range Asymptotic Prioritarianism, and other ways of avoiding the Monstrous Conclusion in population ethics is enormous.

In the next Chapter, we will see that this is a symptom of a more general problem. In any situation where more than a dimension has non-trivial value, despite we may wish to balance all dimensions, a sufficient increase in one dimension can outweigh in importance all other dimensions combined.

Chapter 9: Value Pluralism, Irreducibility, and Non-triviality of Values

Abstract. Value Pluralists hold that there are multiple fundamental values. It has been suggested that, to escape the impasse of the impossibility theorems, we should appeal to Value Pluralism. In this Chapter I argue against this suggestion: a Pluralist theory of population ethics will either be unacceptably hyposensitive or be prone to have only one value “swallowing” all other values – in the case of population ethics, the value in question is welfare. In particular, using what makes a life better or worse as a case study, I show that Value Pluralists are susceptible to an inductive argument that forces them to choose between two principles they are deeply committed to. First, Irreducibility, according to which there are cases where a sufficient amount of many values is better than any amount of just one value. Second, Non-triviality, according to which values have non-trivial weight relative to other values – adding a sufficient amount of one kind of value is better than adding a tiny amount of another value. These principles are central to Value Pluralists, but incompatible: one must be rejected. I argue that Value Pluralists should reject Irreducibility and retain Non-triviality. In the case study of prudential values, some object that pleasure is of trivial value. I respond by showing that the same argument structure can use pain rather than pleasure. No-one could plausibly defend that pain has trivial disvalue. Irreducibility has a problem: it is implausibly insensitive to some values – it is Hyposensitive. This generalises to Value Pluralist theories beyond the case study of wellbeing theories. In conclusion, Value Pluralists should reject Irreducibility and avoid Hyposensitivity. I show that this implies that pluralism can provide no unique, novel, costless way out of the dilemmas of population ethics.

1. Introduction

According to Value Pluralism, there is a multiplicity of values that cannot be reduced to a single overarching value. The Value Pluralist tradition is influential in many different domains: what makes an outcome better or worse, what makes an action politically justified, what makes an action morally permissible or obligatory, and what makes life go best, to name just a few (Chang, 2002, 2015b; Hedden & Muñoz, 2023; Nagel, 2009; Rawls, 2003; Ross & Stratton-Lake, 2002; Temkin, 2012).

A Value Pluralist theory about populations may consider more than the two variables that are classically taken into account, namely, the number of people and their welfare. Population Ethics mostly focuses on the number of people, and their welfare. These are the variables considered in the impossibility theorems, and these are the main variables considered in the Monstrous Conclusion. Thus some may suspect that, by considering more values, as Value Pluralists may, we may be able to escape the impossibility theorems and avoid the Monstrous Conclusion in some novel way, that requires no intellectual cost.

The goal of this Chapter is to show that this Value Pluralist route out of the dilemmas of population ethics is unpromising. Adding more values will simply reproduce the same dimensional problems over and over again, at least so long as these values come in degrees.

There are only a few attempts of Value Pluralist Theories (see (Arrhenius, 2000b, pp. 104-113,139-150, Forthcoming; Feldman, 1995; Temkin, 2012, pp. 161–197)), so I analyse Value Pluralism in a field that presents similar characteristics as population axiology: wellbeing theories. As in population axiology, wellbeing theories presents values that come in degrees, as the “smooth continuum” assumption is as hard to deny in population axiology as it is in wellbeing theory. As in population axiology, there is a tradition of considering a theory problematic if it implies that one dimension of good can trump all others: in population ethics, it is a problem if total quality trumps any other consideration, as in the Repugnant Conclusion (Chapter 3. The Benign Addition Paradox , Chapter 5: Totalism: Between Repugnance and Greediness), or when a single individual’s wellbeing

trumps any other consideration, as in the Monstrous Conclusion (Chapter 8). In wellbeing theories, it is a problem when pleasure trumps any other consideration: as we are about to see, this happens with Robert Nozick's "experience machine" (Nozick 2012, p. 42-43).

Thus, in this Chapter, I am going to focus on wellbeing as a case study, one of the domains where Value Pluralism is most widely endorsed. According to Value Pluralists about wellbeing, many goods intrinsically make one's life better: in our lives we want meaningful relationships, we want to achieve fulfilling projects, we want to be happy, we want to be healthy, have pleasurable feelings and much more. These many different things make a life better in distinctive ways. These values are irreducible to one another and have non-trivial worth relative to one another. I call the first feature Irreducibility, and the second Non-triviality. In this introductory Section, I proceed to introduce them in turn, and then sketch how I demonstrate that they are incompatible.

Firstly, let us examine Irreducibility. While there are almost as many accounts of irreducibility as there are Value Pluralists, all these different accounts agree that the following toy reduction argument fails. Suppose we have a model non-pluralist utilitarian account, which claims that more utils is always better. Each of the good things in our life, projects, relationships, pleasures and so on, are each worth a given number of utils. Therefore, any rich fulfilling life is worth a fixed number of utils, the sum of each of these components. As a unit of pleasure is worth a fixed number of utils, there is some number of units of pleasure which is worth more utils in total than the rich fulfilling life. Hence, according to this model utilitarian account, the life with sufficient units of pleasure is at least as good as the rich fulfilling life. However, as Value Pluralists argue, this conclusion is implausible (Kagan, 1994; Nozick, 2012; Scanlon, 2000; Smart & Williams, 2008; Sobel, 2002; Sumner, 1996; Tiberius, 2013).⁸²

⁸² Note that the argument works even if one does not interpret irreducibility as a notion about how values can be traded off against each other. Some Pluralists think that irreducibility simply means that there is no single property of "goodness" in common among valuable things, and each value makes things good or bad in its own unique way. For example, there is no commonly shared property of goodness between the pleasure of eating good food and the entertainment of listening to a nice symphony, despite both of them being good for us (Parfit 1984, p. 493; for similar claims, see (Brennan, 2006;

This intuition is at the core of Nozick’s “experience machine” argument (2012), a paradigm example of an argument against reducibility, and one of the most influential arguments for Value Pluralism about wellbeing (Sumner 1996; Scanlon 2000; Kagan 1994; Sobel 2022; Tiberius 2013, Crisp 2005). Nozick argues that, even if there were an experience machine that would give one any experience that one desired, we would have a strong intuition against plugging in to the machine. This argument defends what I call Irreducibility: that some lives are better than any life reduced to containing only one value, such as pleasure.

Secondly, consider Non-triviality, according to which values always have non-trivial worth relative to one another. This means that, for any amount of one given kind of value, there is a sufficient amount of a second kind of value such that adding the second value is better than adding more of the first. With this, Pluralists about wellbeing aim to capture the intuition that increases in some values are not worthless if compared to increases in other values.

Some Pluralists have claimed to reject Non-triviality, by stating, for example, that liberty has “an absolute weight, so to speak, with respect to [other goods such as utility and prosperity], and hold without exception” (Rawls, 1972 p. 38; see also p. 36-40; for similar claims, see for example Crisp 2005, Dorsey, 2009)). However, the rejection of Non-triviality is considered radical. Indeed, so radical that, as noted by (Griffin, 2002, p. 83; Nebel, 2022, p. 202) not even their proponents seem to mean them genuinely. For example Rawls himself thinks that an obsolete weight “cannot be strictly correct” (Rawls, 2003, p. 40). See also (Rawls 1972, sects. 26 and 82, esp. pp. 152, 542–3).

Many pluralists accept Non-triviality (Chang, 2016; Griffin, 2002; Hurka, 1987; Lin, 2016; Temkin, 2012, and others). For example, they say that it is at least as good (that is, either better or equally good) to suffer a minor pinprick pain to satisfy one’s life project, because projects are not worthless if compared to pain avoidance. Or, it is at least as good to incur a tiny loss in terms of

Dancy, 2004; Scanlon, 2000; Thomson, 1997; Williams, 1981)). This is a claim about what value fundamentally is rather than a claim about trade-offs. However, this claim about fundamentality implies that we cannot trade all values for a single value.

friendship or knowledge for the sake of a major improvement in one's hedonic state, because pleasure is not worthless if compared to friendship or knowledge. This is because, intuitively, whatever a life is like, it is worth slightly decreasing one of its values to greatly increase another value. I will refer to this widely endorsed intuition as Non-Triviality.

Irreducibility and Non-Triviality are two core features of Pluralism about wellbeing. However, I demonstrate that the two are incompatible. I use a mathematical induction argument – these consist of two parts (this is a form of deductive argument not to be confused with inductive reasoning).⁸³ First, I show that a property applies to a base case and, second, that for an arbitrarily chosen member within a sequence, if the property holds for the given member of the sequence, then it holds for the next member of the sequence. A common example for visualising how an inductive argument works are falling dominoes. Suppose we see a line of dominoes, and we possess two pieces of information about it. First, the first domino will fall: that is our “base step”. Second, when any domino falls, the next domino in the line will fall: that is our “inductive step”. With this information, we can conclude that all dominoes in the line will fall, including the last one.

Inductive arguments have been insightful in various areas of philosophy (Beckstead & Thomas, 2023; Campbell & Mosquera, 2020; Parfit, 1986; Temkin, 2012; Wilkinson, 2022). However, no-one has considered the implications of iterated applications of Non-Triviality for Irreducibility, and ultimately for Value Pluralism.⁸⁴ In Section 3, I show that an inductive argument forces Value Pluralists to choose between Irreducibility and Non-Triviality. In Section 4 I respond to a central objection to the argument given in 3. According to the objection, Non-Triviality is not as compelling as it looks. I present another inductive argument, where I show that Irreducibility is incompatible with a particularly compelling version of Non-Triviality, according to which at least pain can never be trivial.

⁸³ Inductive reasoning consists in deriving a general claim from a series of particular claims. Despite the name, mathematical induction is a *deductive* claim, that is, it derives a particular conclusion from a set of premises.

⁸⁴ See ((Pummer, 2017) for the most similar project.

Finally in Section 5 I show that, these arguments reveal a deeper desideratum for Irreducibility – namely Hyposensitivity Avoidance. Hyposensitivity occurs when the overall wellbeing of a life is insufficiently sensitive to large changes in one value. Ultimately, I conclude that Value Pluralists should avoid Hyposensitivity and therefore reject Irreducibility.

2. Irreducibility and Non-Triviality

The presence of multiple irreducible, non-trivial values lies at the core of Value Pluralism. In this Section, I define Irreducibility and Non-Triviality within the context of wellbeing, and I briefly highlight why they are so compelling.

To better understand how compelling this intuition is, consider the following case. Suppose that the relevant values were projects, health, relationships, security, and pleasure over pain. Imagine the life of Bob. Bob's main project consists in his paintings, and he produces one decent painting per month. He has some sight problems, but can see normally with a set of glasses. He has a partner, one child, and two close friends. He has a stable income, but has to carefully budget his expenses to provide for his family. He has some back pain but derives pleasure from eating nice food. Now suppose there are two ways in which Bob's life could change. One is to improve significantly in all values to get a very good, flourishing life; another is to improve immensely in only one value with a loss of other values to get an impoverished life.

Suppose Bob could produce more paintings per month: instead of one, he manages to produce four excellent paintings, getting invited for art exhibitions on a monthly basis and manages to sell some paintings at a high price. His security improves as the earnings from his paintings allow him to quit his job without worrying about his savings or pension. In addition, his health improves: he has time to exercise and get successful laser eye surgery. Bob has more time to devote to his friends and family, enabling him to play a much bigger role in the growth and development of his child, and

would be able to fully support and care for his partner whenever is needed. This change results in a flourishing balanced life, excellent with respect to many values. Call this the *Flourishing Life*.

Alternatively, Bob could enter a machine that stimulates pleasurable experiences. The machine makes him feel immense amounts of pleasure, all day, every day - nothing compares to the absolute bliss of this machine. He forgets about all his worries and discomfort. He stops painting. His relationships with his friends and family deteriorate. He manages to survive thanks to his savings, but must move into a much smaller house, and does not always manage to pay his bills. His eyes and back deteriorate. This change results in an unbalanced life, that is outstanding with respect to only one value, pleasure, but bad with respect to the other values. Call this the *Impoverished Life*.

The strong intuitions that Value Pluralists have is that the Flourishing Life, excellent in many dimensions, is better than the Impoverished Life, where only one value (pleasure) is outstanding but all the others are lacking.

This argument is modelled after an influential argument by Nozick, a *reductio ad absurdum* of the idea that all that is good in a life can be reduced to what one feels “from the inside” (Nozick 2012, p. 42). In particular, as pleasure entirely consists in what one feels “from the inside”, Nozick’s argument is also a *reductio* of the idea that all that is good in a life can be reduced to pleasure. This is Nozick’s argument in a nutshell:

Suppose there were an experience machine that would give you any experience you desired. [...] Should you plug into this machine for life [...] We learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it. (Nozick 2012, p. 42-43).

More explicitly aimed to pleasure, a strikingly similar argument published 13 years earlier by J. J. C. Smart targets the reduction to a life purely driven by pleasure ((Smart & Williams, 2008), p. 19. Do

not let the date of the editions I cite fool you: the original edition of the two texts are, for Nozick's book, 1974, and for Smart's, 1961).

[..] a bald-headed man with a number of electrodes protruding from his skull, one to give the physical pleasure of sex, one of that of eating, one for that of drinking, and so on. Now is this the sort of life that all our ethical planning should culminate in? [...] Surely not. Men were made for higher things, one can't help wanting to say [...]. (Smart & Williams 2008, 19)

These experience machine arguments constitute a major tradition in arguments for Value Pluralism about wellbeing, and one that is widely considered to land a "near-fatal blow" to non-pluralist objectivist theories such as Hedonism ((Crisp, 2006, p. 99). A widely accepted argument against non-pluralist theories is a widely accepted argument in favour of pluralist theories (Sumner 1996; Scanlon 2000; Kagan 1994; Sobel 2022, p. 244; Tiberius 2013, Crisp 2005). The argument, in a nutshell, argues for what I'll call

Irreducibility. Some lives are better than any life reduced to containing only one value, such as pleasure.

Let us now examine Non-Triviality, that is the intuition that some increases in values have nontrivial worth compared to others. Pluralists have been pointed out that "Nearly all of us would sacrifice some liberty to avert a catastrophe, or surrender some autonomy to avoid great pain" (Griffin 2002, p. 83), that "any remotely plausible view [...] would reject the conclusion that a second of extra life would outweigh, say, Northern Lights viewing" (Dorsey 2009, 47), and Nozick rhetorically asks "What's a few moments of distress compared to a lifetime of bliss?" (2012, 43). This is because, whatever a life is like, it is not worth to slightly decrease one of its values to greatly increase another value. Indeed, one may think it would be irrational not to slightly decrease one value to greatly

increase another. Many others consider this point to be beyond any reasonable doubt (Broome, 2004, p. 24-26; Griffin, 2002; Hurka, 1987; Temkin, 1987, 1996, 2012).

In order to understand why Non-Triviality is so compelling, consider Bob again, with his paintings, partner, child, friends, and everyday pleasures and pains. Suppose that Bob could slightly decrease one of his life's values in order to greatly increase another value. For example, suppose that Bob can shorten a dinner with his partner by five minutes in order to have one of his paintings in an exhibition for the first time. This would involve a very small decrease with respect to his relationship, but a great increase with respect to his project. It is at least as good for Bob to have a shorter dinner and the painting at the exhibition (than it is for him to have a longer dinner and no painting at the exhibition). Indeed, it might seem irrational for Bob not to shorten the dinner and to have the painting at the exhibition. Non-Triviality captures this intuition.

An important element of Non-Triviality is that whether there is some big enough gain (of one value) to make it worth incurring a small loss (of another value) does not depend on what the rest of the life is like. It is so clear that the combination of the two changes does not make one's life worse, that it would be perverse to first ask what the rest of one's life involves – such information is simply irrelevant. For example, imagine someone asks us whether developing a deep and meaningful relationship is worse than one more lick of a lollipop. The answer is so clear we would be confused as to why they are asking it at all, and the answer does not depend on what the rest of their life is like. Even if the life contains no pleasure and lots of meaningful relationships, having the deep and meaningful relationship cannot be worse than having a lollipop lick.

I define this intuition as

Non-Triviality (intuitive). Whatever a life is like, it is at least as good (that is, better, or equally good) to slightly decrease one of its values to greatly increase another value.

Non-Triviality (precise). For some fixed slight decrease d in one value, there is some sufficiently great increase I in another value such that, whatever a life is like, it is at least as good (that is, better, or equally good) to slightly decrease one of its values by d to greatly increase another value by I .

Both Non-Triviality and Irreducibility are intuitively compelling. However, they are incompatible.

3. The first argument

A first inductive argument for the incompatibility between Irreducibility and Non-Triviality will have the following shape. Starting with a flourishing life, there is a sequence of changes that, according to an Inductive Step based on Non-Triviality and Transitivity of “at least as good as”, would each produce a life that is at least as good as the previous, but the combined result is an Impoverished Life. A given life is “at least as good as” another life only if the given life is either as good as, or better than, the other life. Transitivity of “at least as good as” says that, if a life A is at least as good as a life B, and a life B is at least as good as a life C, then a life A is at least as good as a life C. As an implication of the Inductive Step, the Impoverished Life at the end of the iterated series of changes is at least as good as the initial flourishing life. However, according to Irreducibility, that directly compares the initial and final life, the flourishing life is necessarily better than the unbalanced life. Therefore, Non-Triviality and Irreducibility are in contradiction.

The argument is based on the following two premises. The first is a truism, so I will not argue for it:

Base Step: a flourishing life is at least as good as itself.

The second premise follows from Non-Triviality and what is known as transitivity of “at least as good as”. Non-Triviality says that, whatever a life is like, it is at least as good to slightly decrease one of its values to greatly increase another value. Transitivity of “at least as good as” says that, if a life A is at least as good as a life B, and a life B is at least as good as a life C, then a life A is at least as good as a life C.

As we have seen in Chapter 4, Transitivity of “at least as good as” is considered a truism by many, and has great support among philosophers (Broome, 2004; Huemer, 2008; Nebel, 2018; Pummer, 2018 and others). There are some important sceptics of *a* transitivity of “at least as good as”, such as Tim Scanlon, Derek Parfit, and Larry Temkin, but that is not the transitivity we will be referring to. They distinguish between a sense of “at least as good as” which tracks how good the internal features of an outcome are, and a sense of “at least as good as” which tracks how preferable an outcome is if compared with another. That is, they are rejecting “teleology”: see Chapter 2. Since these authors think that internal features are not always what makes an outcome at least as good as another, they do not think transitivity of “at least as good as” is always true. However, even these authors agree that, when it comes to the goodness of the internal features of an outcome, Transitivity of “at least as good as” is analytically true (Temkin 2012, ix/10-18). Since this Chapter is concerned only with the internal features of a life, I will assume Transitivity of “at least as good as”.

Together, Transitivity of “at least as good as” and Non-Triviality imply:

Inductive Step: if a given life is at least as good as the flourishing life, then a life that is slightly worse than the given life in one of its values but much better with respect to another value is at least as good as the flourishing life.

I show that these two premises, *Inductive Step* and *Base Step*, are incompatible with Irreducibility, according to which some lives, and particularly flourishing lives, are better than any life reduced to containing only one value, such as pleasure.

To see how the argument runs more clearly, recall Bob's *Flourishing Life*. Bob is a successful painter, a caring partner and parent, is healthy, and is financially secure. This life is the first member of our sequence. This life, our *Base Step* premise says, is at least as good as itself.

Next, consider a similar life in which Bob paints one fewer brushstroke and has six additional hours of pleasure, generated by an experience machine.⁸⁵ Since the pleasure is incredibly intense, the amount of pleasure in his life would increase enormously, at the very small price that one of his paintings will be slightly less refined. Surely forgoing a single brushstroke for the sake of gaining six hours of absolute bliss does not make the life worse than the initial Flourishing Life. This is what the *Inductive Step*, informed by Non-Triviality and transitivity, indicates. Indeed, some may think it is irrational not to forgo the single brushstroke to gain six hours of bliss.

Suppose we repeat choices of this kind. Bob is offered to trade a second brushstroke with six additional hours in the experience machine. This would not be worse for him, given the small price and the great gain. We now keep giving Bob the option to enjoy six hours of absolute bliss at the small cost of removing a single brushstroke on his painting. And again, it would not be worse for Bob to get another six hours of unimaginable bliss at the small price of a third brushstroke less. And so on: if Bob repeatedly chooses hours of pleasure in the machine over brushstrokes, things would never get worse than the Flourishing Life according to the *Inductive Step*. But as the brushstrokes get fewer and fewer, the quality of his paintings deteriorates. Ultimately, Bob's paintings will have so few

⁸⁵ If one finds that 6 hours of pleasure do not compensate the number of brushstrokes, one can increase the number of hours to 12, 18, or more. Also, note that while Bob's project does not *consist in* a series of brushstrokes and Bob's relations do not consist in time spent with people, it is still instructive to consider changes such as one brushstroke at a time or five minutes with people at a time. In fact, by considering such small changes in the conditions of possibility for values such as projects and relationships, we ensure that we are necessarily considering small improvements or worsenings of these values. It is possible that for some brushstrokes, or for some minutes spent with loved ones, there will be no improvement with respect to projects, or relationships. It is also possible that for some brushstrokes, or for some minutes spent with loved ones, there *will* be improvements much bigger than others, but if you take away the minutes or brushstrokes that make no difference, these big improvements will not happen.

strokes that they are of no artistic relevance to anyone: his project concerning his paintings dramatically fails.

Next consider another kind of value in Bob's life. Suppose Bob can spend six additional hours of pleasure if he devotes slightly less time to his loved ones – for example, a meal or a walk with them will last a few minutes less. If a single activity with his loved ones lasts less long, his relationships will not be noticeably affected for the worse: it is at least as good for Bob to shorten a walk with his loved ones in order to enjoy six hours of additional intense pleasure. And if he were offered to shorten the activity even further, or to shorten a different one, in order to enjoy other six hours of pleasure, it would surely not be worse for him to do so. If he were to receive a similar offer a third, fourth, fifth time, and so on, on each occasion, it would not be worse for him to accept. But then, he would end up spending little to no time with his loved ones, and his relationships would be very shallow and unsatisfying. And yet, given the *Inductive Step*, things are never getting worse than the *Flourishing Life*.

Similarly, Bob will be in the position to give up little by little all the remaining values, whatever they may be (in our example, health and financial security), to gain enormous pleasure at each time. Intuitively, whatever the life is like, very small quantities of one value cannot be worth more than great quantities of any other value. Six hours of intense pleasure in the machine is a great amount of pleasure. For each choice and each value, it is at least as good to sacrifice the very small amount of that other value for six more hours of intense pleasure. Thus, again, given the *Inductive Step*, things are never getting worse than the *Flourishing Life*.⁸⁶

Each accepted trade has not been for the worse, but accepting all trades results in the *Impoverished Life*. Bob's initial life, the *Flourishing Life*, was excellent with respect to many values: his project as a painter was successful, he had a great role in caring for, supporting and growing his family, he was in very good health and so on. Then, Bob underwent a series of steps where he gained

⁸⁶ Some may worry that Bob's life duration must be extraordinarily long to ensure enough pleasure to compensate all the steps. While this makes the example unrealistic, pluralists clearly state that their theory is supposed to be able to accommodate any logically possible life (Temkin 1996 p. 185-186; 2012, chapter 5, Lin forthcoming, section 1).

a lot of pleasure in a machine at a trivial cost with respect to other values. The final life is poor in all respects, except pleasure.

According to the *Inductive Step*, no change in the sequence is worse than the initial *Flourishing Life*. Thus, the last *Impoverished Life* is at least as good as the first *Flourishing Life*. By contrast, Irreducibility says that the first, *Flourishing Life*, very good with respect to many values, is better than the last *Impoverished Life*, that is outstanding with respect to only one value but poor with respect to other values.⁸⁷

We have a contradiction between Non-Triviality and Irreducibility. Value Pluralism about wellbeing cannot rest on contradicting principles. One of the two principles has to be given up. Which one should we reject? Which condition is least plausible, and should be abandoned? The remainder of this Chapter examines how we should respond to this contradiction.

Before continuing, however, it is worth noting what follows. (Binmore & Voorhoever 2003) argue that inductive arguments of the kind I propose here, and Temkin (1996) and Parfit (1986) have proposed in the past, do not work. This is because these arguments do not exclude that any finite sequence of slight decreases in on one value may never take us all the way to the absence of that value, but instead asymptotically converges to some positive level of this value. However, Carlson (2005) has shown that, so long as the first step is formulated with fixed values, as I do with *Non-Triviality (precise)*, then Broome and Voorhoever's objection does not work. Additionally, Temkin argues that the presence of a finite sequence that goes from any degree of realization of a value to its absence is also sufficient to make the argument work (2012, appendix D2). While Broome, in personal conversation, tells me he finds it implausible that there is a fixed decrease in one value that brings it from any degree of realization to its absence, Voorhoever has stated in print that he considers

⁸⁷ Some may think that the *Flourishing Life* may not be the kind of life that would be better than the Unbalanced Life according to Irreducibility. If this were true, this would be a problem for Pluralism, since for any life that is better than the *Unbalanced Life* according to Irreducibility, there is an *even better* life that is excellent with respect to many values, but is worse than the *Unbalanced Life*. This violates transitivity of "at least as good as".

Carlson's reply is satisfactory (Voorhoever 2008, note 4). I modelled *Non-Triviality (precise)* after Carlson's paper, and I side with him Voorhoever, but against Broome, in thinking that we *can* decrease any value from any degree of realization to its absence by a fixed amount.

4. The second argument

The inductive argument in the previous Section I highlighted the incompatibility between Non-Triviality and Irreducibility. Pluralists have to give up one of the two intuitions. I expect that a pluralists' reaction to this inductive argument is to consider Non-Triviality less compelling than Irreducibility. Irreducibility is a core feature of pluralist theories, capturing the multiplicity of the nature of value. On the other hand, Non-Triviality is pointing out specific applications of pluralist wellbeing theories, that may be peripheral to the central commitments of these theories. It may be that Non-Triviality, despite looking plausible at first glance, commits us to more than initially recognised if applied over and over again.

However, before rejecting Non-Triviality in favour of Irreducibility, we ought to investigate alternative variations of the iterative argument. In fact, if Irreducibility is incompatible with premises even more compelling than Non-Triviality, Irreducibility will need to be given up, and Non-Triviality can be retained.

In this Section I propose a second inductive argument, which involves an iterative premise which is similar to Non-Triviality but much harder to reject. Consider

Pain Non-Triviality (intuitive). Whatever a life is like, it is at least as bad (that is, worse, or equally good) to slightly increase one of its values to greatly increase the amount of pain in this life.

Pain Non-Triviality (precise). For some fixed slight increase d in one value, there is some sufficiently great increase I in pain such that, whatever a life is like, it is at least as bad (that

is, worse, or equally good) to slightly increase one of its values by d to greatly increase the amount of pain in one's life by I .

While Non-Triviality says that some increases in value have non-trivial worth if compared to others, Pain-Non-Triviality stresses that an increase in pain has non-trivial (negative) worth if compared to an increase of other values. For example, whatever Bob's life is like, an additional paint brush on one of Bob's canvases is not worth hours of back pain, and it is at least as bad to put a paint brush on the canvas if this means suffer back pain for hours. Indeed, one may want to say that it would be irrational to brush if this causes so much pain.

Non-Triviality is a hard premise to reject but Pain-Non-Triviality is a significantly harder to reject premise. To see this, suppose Bob was having dinner with his partner and chooses to leave five minutes earlier than planned. Bob's partner asks him why he is leaving. Consider two scenarios.

In the first scenario, Bob answers "to experience intense pleasure in the experience machine." Bob's partner replies, "you've spent all week in the experience machine, please prioritise having our meal together next time?" Bob replies, "Why does the amount of time I've already spent in the machine matter? Do you not care about the vast amount of pleasure I would have to forego?". In this scenario, while it is plausible that it is better for Bob to cut the meal short, Bob's partner also comes across as entirely reasonable and Bob as insensitive. There is at least something to Bob's partner's point that Bob has already spent a considerable time in the experience machine at the cost to their relationship. It is not clear whether this sensitivity to what the rest of what Bob's life is like is about prudential value rather than some other deontic consideration, such as fulfilling promises, but there is some scope for doubting Non-Triviality in this case.

In the second scenario, Bob answers "I've been experiencing back pain and I need to rest it to avoid experiencing hours of intense pain." Bob's partner replies, "you've spent all week resting your back, please prioritise having our meal together next time?" Bob replies, "Why does the amount of time I've already spent resting matter? Do you not care about the vast amount of pain I would have

to endure?” In contrast to the first scenario, now it is Bob’s partner who comes across as unreasonable and insensitive about what is good for Bob. This time it doesn’t seem to matter how long Bob has been resting, the extent of the pain that he would face means that it is totally reasonable for him to shorten the meal regardless of what the rest of his life is like – Pain-Non-Triviality clearly applies. The significance of the badness of pain for Bob overrides whatever explanation there may be in the first scenario, so Pain-Non-Triviality is a much harder to reject premise than Non-Triviality. Having established that Pain-Non-Triviality is even harder to reject than Non-Triviality, I now present an inductive argument showing that Pain-Non-Triviality is incompatible with Irreducibility.

The argument is based on the following two premises. As for the previous inductive argument, the first premise is a truism, so I will not argue for it:

Base Step (II): an Impoverished Life is at least as bad as itself.

If given life is “at least as bad” as another, then the given life is either worse than the other life, or equally good. The relation “at least as bad as” is transitive: if a life A is at least as bad as a life B, and a life B is at least as bad as a life C, then a life A is at least as bad as a life C. I defended transitivity of “at least as good as” for the previous argument: the defence of transitivity of “at least as bad as” is identical, and I shall not repeat it.

The second step follows from Pain-Non-Triviality and transitivity of “at least as bad as”. Pain-Non-Triviality says that, whatever a life is like, it is at least as bad to slightly increase one of its values to greatly increase pain. Together, transitivity of “at least as bad as” and Pain-Non-Triviality imply:

Inductive Step (II): if a given life is at least as bad as the Impoverished Life, then a life that is slightly better than the given life in one of its values but much worse with respect to pain is at least as bad as the Impoverished Life.

I show that these two premises are incompatible with Irreducibility, according to which some lives, and particularly Flourishing Lives, are better than any life reduced to containing only one value, such as Impoverished Lives.

Suppose that Bob is starting from the *Impoverished Life*, where he experiences a lot of pleasure in the machine, but is miserable in all other values. From this life, Bob could make a sequence of choices. In each choice, he will suffer back pain for six of the hours he is in the experience machine, but slightly improve one of the values that is not pleasure. Improving all the values little by little, his life will become excellent with respect to all values, all at the price of substantial back pain.

To avoid the worry that the back pain makes Bob's life unbalanced, we can assume that Bob's back pain is immediately largely compensated by the pleasure he experiences in the machine.⁸⁸ Because of the intense pleasure of the experience machine, Bob prefers to be conscious for the six hours in the experience machine with the back pain rather than taking a sedative so Bob would not be conscious while his back recovers.⁸⁹ Since the pleasure he experiences in the machine greatly compensates for the back pain and, given all the pleasure generated by the machine, Bob's life is excellent with respect to pleasure over pain.⁹⁰

According to the inductive step, no choice makes Bob's life better than the *Impoverished Life*. However, by repeatedly choosing to make small improvements at the cost of great (compensated)

⁸⁸ Some may think that compensated back pain is very different from the kind of back pain Bob is feeling if he stays at dinner with his partner longer in the previous thought experiment. One can assume Bob would enter the experience machine shortly after dinner anyway, and this compensates the pain. Denying that pain can be compensated by events happening shortly after would imply to the absurd conclusion that, if the machine had one second of pain and then six hours of pleasure shortly after, this would be worse than having constant pain and pleasure for six hours.

⁸⁹ Consider the experience of having a purring cat on your lap – this experience is intensely pleasurable. Now suppose the cat is padding your leg, extending and retracting its claws each time. The experience is simultaneously pleasurable and painful, where the relative amounts of pleasure and pain determine whether the experience is good, neutral or bad overall. Alternatively, if one is committed to the claim that pleasure and pain cannot co-exist, consider a variation where pleasure and pain rapidly alternate where Bob overall is prefers to have more time in the machine, is indifferent to more time, or prefers not to have more time in the machine respectively.

⁹⁰ I am assuming that enough pleasure can compensate back pain. I need not to assume that any pain, no matter how intense, can be compensated.

pain, Bob eventually reaches a *New Flourishing Life*, that is very good with respect to all values. Since, according to the inductive step, no life in the series of modification starting from the *Impoverished Life* and finishing with the *New Flourishing Life* is better than the *Impoverished Life*, the *New Flourishing Life* is *at least as bad as* the *Unbalanced Life*. This contradicts Irreducibility, according to which a *Flourishing Life* good in all dimensions, *is* better than any “Impoverished Life”, good only with respect to one value.

To see how the argument runs, consider Bob’s *Impoverished Life*. Bob feels intense pleasure every day, all day for an incredibly long life but with no other sources of value. He does not paint nor has exhibitions; he does not care about his family; he struggles to pay his bills and has a bad back. His life is outstanding with respect to only one value, pleasure, but bad with respect to the other values. By *Base Step (II)*, this life is at least as bad as itself. I take this to be uncontroversial.

Suppose Bob could put a single brushstroke on a canvas at the cost of experiencing six hours of back pain when in the experience machine. This would be a trivially small improvement with respect to Bob’s project, not worth more than substantial suffering: according to the *Inductive Step (II)* supported by Pain-Non-Triviality and transitivity of “at least as bad as”, this life is at least as bad as the *Impoverished Life*. Indeed, one may want to say they are getting worse.

Suppose we keep offering Bob choices of that kind. Suppose that Bob could put on a canvas a second brushstroke, but in doing so, he would suffer a second six hours of back pain. This is another very small improvement with respect with projects, that is not worth more than substantial suffering: according to the *Inductive Step (II)*, Bob’s life is not getting better than the *Unbalanced Life*. Indeed, some may want to say it’s getting worse, and it is irrational for him to put the brushstroke in the canvas. And again, it would not make Bob’s life better to trade yet a third period of back pain for adding a third brushstroke and so on. But as the brushstrokes get more and more, the quality of Bob’s paintings keeps on improving. Ultimately, Bob’s painting will be incredibly refined. Indeed, so refined that he’ll produce four excellent paintings per month, sells some of them at a high price and is invited to exhibit monthly: his project concerning his paintings would greatly flourish.

Next consider another value Bob could improve in his life. Suppose Bob could devote five more minutes to his loved ones if, for another six hours in the experience machine, Bob experiences back pain. For example, in these five minutes with his family, he can participate in very small pieces of walks, or meals, with his family. If he gets just five minutes to interact with his loved ones, his (currently non-existent) relationship with them will be affected for the better only very little. According to the *Inductive Step (II)*, it is at least as bad for Bob to going through six hours of back pain for such a small gain do not make Bob's life better than the initial *Impoverished Life*. Indeed, some may say it would be worse, and it would be irrational to interact with his loved ones for so little time at such a high cost. And if he received a similar offer a second, third, fourth, fifth time, and so on, according to the *Inductive Step (II)* accepting it would never be for the better – some may want to say that it would be worse and worse. But, if Bob keeps accepting them, he would reach a large time to devote to his friends and family, enabling him to play a much bigger role in the growth and development of his child, and would be able to fully support and care for his partner whenever is needed: his relationships would be excellent.

Bob will be offered similar choices for the remaining values, whatever they may be (in our example, health and financial security). The final outcome is that for some finite number of additional hours of back pain, Bob will have a flourishing life in many respects, will produce four excellent paintings a month, and get invited for art exhibitions. The earnings from selling his paintings means he does not need to worry about his savings or pension or how to look after his family, would be able to fully support and care for his partner and so on. In addition to these aspects of his life, Bob also experiences a long period in the experience machine involving pleasure and back pain, but the pleasure is so much greater than the pain, he prefers to experience it than to be sedate for the pain.⁹¹ This change results in a flourishing balanced life, excellent with respect to many values. Consequently, according to Irreducibility, this life is better than the original life, which was only

⁹¹ The difference between having an experience which one equally prefers and is only good with respect to hedonistic value may have some affect but can't be so much as to violate the ranking of lives here.

valuable with respect to pleasure. But this is incompatible with the inductive premise, according to which no life in the sequence is better than the initial *Impoverished Life*.

To summarise this Section, starting with an unbalanced life with one value outstandingly high and others very low, by making a series of changes that are at least as bad according to an *Inductive Step (II)* based on Pain-Non-Triviality and transitivity of “at least as bad as”, we end up with a flourishing life which, by Irreducibility, is better than the initial *Impoverished Life*. Pain-Non-Triviality and Irreducibility are in contradiction. Given that Pain-Non-Triviality is much harder to reject than Non-Triviality, the case for rejecting Irreducibility is significantly stronger.

5. Hypersensitivity, Hyposensitivity, and Trumping

The case for rejecting Irreducibility can be further strengthened by learning a more general lesson from the way in which Non-Triviality and Pain-Non-Triviality conflict with Irreducibility. While Non-Triviality and Pain-Non-Triviality are quite specific principles, they illuminate more fundamental ways in which Irreducibility conflicts with central claims about the relative worth of values and plausible ranges of sensitivity between values.

The different values that have been considered in this Chapter have non-trivial worth with respect to one another. For example, it would be better for me, given my life, to suffer a pinprick pain to satisfy my life projects or to miss one minute of a meal with a friend to avoid significant amounts of pain.

By contrast, some people have claimed that certain prudential goods or bads “trump” others, that is, some goods or bads are worth more than any amount of some others (the term “trumping” is by Griffin, 2002, p.83. These goods “trumping” others are the Superior goods of Chapter 3). For example, Ross (Ross & Stratton-Lake, 2002) claimed that any amount of virtue, no matter how small, is worth more than any amount of pleasure, no matter how great. As he puts it “[w]ith respect to pleasure and virtue, it seems to me much more likely to be the truth that no amount of pleasure is

equal to any amount of virtue, that in fact virtue belongs to a higher order of value, beginning at a point higher on the scale of value than pleasure ever reaches...” (Ross & Stratton-Lake, 2002, p. 150).⁹² Similarly, Rawls says that the principle of ensuring liberty has “an absolute weight, so to speak, with respect to [other goods such as utility and prosperity], and hold without exception” (Rawls 1972 p. 38; see also p. 36-40).

While there have been a number of famous claims concerning trumping values, these claims are implausibly strong. Indeed so strong that, as mentioned in the introduction, it has been suggested that not even their proponents mean them genuinely. As Griffin reports (Griffin, 2002, p. 83), Rawls himself admits thinking that liberty has priority only once after a certain level of wealth has been attained (Rawls 1972, sects. 26 and 82, esp. pp. 152, 542–3). The difference between trumping and non-trumping goods is a radical difference in orders of value.

One way of capturing the implausibility of such claims, when they are false, is that they are either Hypersensitive or Hyposensitive (Beckstead & Thomas, 2023; Campbell & Mosquera, 2020; Pummer, 2018, 2022; Thornley, 2021; Wilkinson, 2022). *Hypersensitivity* is a combination of ‘hyper’, meaning ‘too much’ or ‘excessive’, and ‘sensitivity’, meaning how one thing changes as another changes; *Hyposensitivity*, by contrast, is when one thing is implausibly insensitive to another. Ross’ claim about the relationship between virtue and pleasure must either be: Hyposensitive with respect how valuable pleasure is (insufficiently sensitive to massive increases in pleasure); Hypersensitive with respect to how valuable changes in virtue are (so sensitive to arbitrarily small changes in virtue that they outweigh the significant amount of value generated by enormous amounts of pleasure); or both.

A more popular position than “trumping” values is ‘Weak Superiority’, according to which a sufficient amount of one kind or set of values is better than any amount of another kind of value (Arrhenius, 2005; Arrhenius & Rabinowicz, 2005, 2015; Dorsey, 2009 p. 40; Griffin, 2002 p. 85;

⁹² While Ross is talking about pluralism about moral value and not talking about pluralism about prudential value, the point applies equally to both kinds of pluralism.

Parfit, 2016). This does not directly imply sensitivity issues as it need not be small amounts of the superior value which is better than any amount of the inferior value. It certainly seems more plausible to say that a sufficient (potentially quite large) amount of virtue may be better than any amount of pleasure, than that any amount of virtue is better than any amount of pleasure.

Irreducibility captures an instance of Weak Superiority. It says that a sufficiently great amount (potentially quite large) of many values is better than any amount of just a single value. In Irreducibility, what is ‘weakly superior’ is having many values above a high enough threshold, and what is ‘weakly inferior’ is having only a single very high value, no matter how high.

However, as illustrated by the arguments from Non-Triviality and Pain-Non-Triviality, ‘Weak Superiority’ implies Hypersensitivity issues when one is considering a series of options linked by an “inductive step”, and is assuming transitivity (see also Chapter 3). Large amounts of the Superior value can be broken down into very small pieces (possibly arbitrarily small, so long as they are all of the same fixed size), and a series of gains or losses of these small pieces of Superior value is equivalent to gains or losses involving large amounts of this Superior value.⁹³ Consequently, if one claims that a sufficient amount of one kind or set of values is better than any amount of another kind of values, one must also claim that there are cases where an arbitrarily small amount of one value is better than any amount of the other kind of value – there must be a ‘discontinuity’ (Griffin 2002, p. 85-89) where one value “trumps” the other.

Such cases are implausible in the same kind of way that Ross’ claim is implausible. Views of this kind entail that there are cases where an arbitrarily small amount of one value is better than any amount of the other value. While we may not know exactly what are the boundaries within which the sensitivity to e.g. virtue is acceptable and beyond which we have Hypersensitivity, there must be

⁹³ Some may deny that there are arbitrarily small pieces of some goods (Scanlon 2000, pg. 235, Brennan 2006). For example, projects and relationships may not be divided in arbitrarily small units. However, as I showed with the Bob cases, projects and relationships still develop in small bits. And the difference between any two bits is, if not arbitrarily small, at least very small. Indeed, sufficiently small that it is hard to believe that one of these improvements in bits can ever be better for one’s wellbeing than an arbitrarily improvements in another value.

some degree of sensitivity to an arbitrarily small amount of virtue which is too much compared to the large amount of value produced by an enormous amount of pleasure.⁹⁴

Additionally, in some ways the Weak Superiority position is even harder to defend than Ross' initial claim. Whatever defence one might be able to give for why virtue really is in another order of value altogether than pleasure, and thus virtue trumps pleasure, this defence must be false according to Weak Superiority accounts. In fact, in Weak Superiority accounts, when the quantities of pleasure and virtue are low, these values have non-trivial relative worth to one another, and there is no trumping between them. Moreover, so long as an appropriate amount of pleasure is chosen, there will be a sequence where, at one step of the sequence, it is at least as good to gain that amount of pleasure by losing the fixed small amount of virtue but, for the very next trade, not only is losing the small amount of virtue worse than gaining the great amount of pleasure, but suddenly pleasure has trivial value if compared to virtue, and virtue now trumps pleasure.

Given how continuous the variations in the respective values are, such a radical change in the value relation is an implausibly extreme claim. Weak Superiority views must imply this implausibly extreme claim: there are many cases where the two values have non-trivial worth if compared to one another; there are many cases where one of the values has trivial value if compared to the other, and can be "trumped" – a radically different situation; *and* that the difference between the two kinds of cases can be arbitrarily small. The outlook for such a defence is bleak and such views entail implausible sensitivity between the two values in these steps in the sequence.

Irreducibility implies at least a kind of Weak Superiority, as it holds that a life which is at least sufficiently (very) good in many values is better than any life which is outstanding (no matter how good), with respect to just one value. Given the implausibility of complete "trumping" and the arguably worse position for Weak Superiority views, Value Pluralism about wellbeing cannot be sustained.

⁹⁴ To deny that enormous amounts of pleasure produces a significant amount of value would just be to fall into the trap of Hyposensitivity.

To give another perspective on how Irreducibility gives either too much sensitivity on some values or too little to others, consider two ways of expressing Value Pluralism. On what seems to be the standard way to model Balanced Pluralism, if a life is not flourishing, there is a limit to how much each value can improve how good a life is, while if a life is flourishing, there is no such limit. On this view, an improvement of values in a non-flourishing life does make a life better, but there are diminishing marginal increases of value, and an asymptotic limit on how much better they can make a life. This view is *hyposensitive*, since increasing a value close to the asymptotic limit of non-flourishing lives matters too little. In fact, an enormous increase in one value very close to the asymptotical limit in a non-flourishing life is worth less than a slight improvement in the same value in a flourishing life.

Alternatively, one may think a sufficient number of values (or the right kind of values) reaching some sufficiently high threshold make a life better than any life with enough values that are below some threshold. To my knowledge such a bizarre view has never been defended, but it avoids hyposensitivity issues, at the cost of hypersensitivity. In fact, this view is too sensitive to changes around the threshold. Consider a value that slightly increases, but the slight increase takes the value just above the relevant high threshold, and transforms this life from a life that is not flourishing to a life that is flourishing. This slight increase makes that life better than any great increase of another value could.

To summarize, Value Pluralism about wellbeing can't avoid both Hypersensitivity and Hyposensitivity. Avoiding Hypersensitivity and Hyposensitivity is a basic requirement for any value theory: it prescribes that there cannot be excessive sensitivity or implausible insensitivity for kinds of value. Ultimately Value Pluralism, while initially attractive, must be rejected.

While in this Chapter I focused on Value Pluralism about wellbeing as a case study, my arguments generalise. The general structure of Value Pluralism about wellbeing, based on Irreducibility and Non-Triviality, seems to be present in any domain where philosophers have tried to apply Value Pluralism. Indeed, Value Pluralist theories are often presented as applying with an

identical structure across different domains, where wellbeing is just one among many, if not even a paradigmatic case (Chang, 2002, 2015b, 2016; Hedden & Muñoz, 2023; Nagel, 2009; Temkin, 1987, 1996, 2012 and others). Since Irreducibility and Non-Triviality are typically present in all these models, the arguments of this Chapter are likely to apply to all domains where value pluralism applies.

6. Conclusion

Irreducibility and Non-Triviality of different values are central commitments of many who defend Value Pluralism about wellbeing. Irreducibility says that there are lives that are better than all lives that only consist in pleasure. Non-Triviality says that no value has trivial worth compared to another value. In this Chapter, I have showed that Irreducibility and Non-Triviality are compatible. Even if one objects to Non-Triviality, there is a particularly compelling version of Non-Triviality, according to which suffering does not have trivial worth compared to other values, that is also incompatible with Irreducibility.

This incompatibility reveals that Irreducibility necessarily assigns either too much importance to small differences in some values, which is a problem known as Hypersensitivity, or too little importance to great differences in some other values, which is a problem that I call Hyposensitivity. No value theory can forego correctly comparing differences in value. Therefore, Value Pluralists should reject Irreducibility.

If Value Pluralists reject Irreducibility, then their theory implies that one dimension of good can trump all others, and that any considerations about the many values can be outweighed by a sufficient increase in one single value. This means that simply adding value dimensions is not a way to avoid the Repugnant Conclusion, where total welfare is trumping all other dimensions, the Monstrous Conclusion, where an individual's welfare is trumping all other values, or Nozick's experience machine.

Note that accepting Hypersensitivity or Hyposensitivity would not be neither a costless solution to the dilemmas of population ethics, nor a novel one. As we have seen in Chapter 3. The Benign Addition Paradox , Hypersensitivity and Hyposensitivity avoidance are linked to violations of Non-Anti-Egalitarianism. And, the Absolute Priority Condition just is Hyposensitivity. Thus, Value Pluralism offers no unique, novel, costless way out of the dilemmas of population ethics.

Chapter 10. Conclusion

Sailors and navigators in the Ancient Greek world were recommended to travel by sea only within some specific limits. In particular, they were warned not to cross the Pillars of Hercules, which roughly corresponds to the Strait of Gibraltar. According to Greek mythology, the Pillars of Hercules were built by the hero Hercules, during the completion of his “twelve labours”, at the westernmost point where he travelled. Beyond the Pillars of Hercules, dangers of all sorts were said to await the sailors, in the form of monsters and extreme climate events. The warning against crossing the Pillars of Hercules is documented to have survived into the Middle Ages, where the warning even reached Christopher Columbus, who ventured beyond them regardless (Winsor, 2017, p. 128).

The Ancient Greek sailors, however, had little reason to sail across the Pillars of Hercules. For these sailors there was no economic interest, no threat from invaders, no desire for conquest strong enough to brave the sea beyond the Pillars, where monsters and extreme climate events awaited.

Like the pillars built by Hercules which discerned safe waters from unsafe waters, some impossibility theorems, such as the ones in population ethics, discern where moral theorising may be safe from counterintuitive implications, and where instead moral theories are assured to violate at least one compelling adequacy condition (see Chapter 3. The Benign Addition Paradox). However, unlike Ancient Greek sailors, we have no option but to venture into the unsafe waters of population ethics. This is because, despite the sea of population ethics being crowded by monsters, there are decisions that must be taken where the number of people and their wellbeing varies.

The most pressing needs come from climate change. Our policies to deal with climate change are likely to impact the number of people and their wellbeing. Without knowing what is the best way to impact these variables, we cannot propose climate policies in a fully responsible manner (see Chapter 1. Introduction: Aggregating Value).

Philosophers have proposed several ingenious theories to venture into the dangerous seas of population ethics.⁹⁵ In this thesis, I showed novel and serious problems for some of these theories. In particular, I have argued in Chapter 4: Transitivity of “better than” and inferences that theories violating transitivity of “better than”, or theories presenting non-trichotomy that is not restricted to a range, are incapable of drawing seemingly obvious inferences in cases of partial information. Furthermore, I have argued for two things in Chapter 9. First, that Value Pluralist theories cannot provide any unique, novel, costless solution to the challenges of population ethics. Second, that these theories cannot avoid a fundamental tension between their two core commitments. Since Value Pluralism is used in many ethical domains beyond population ethics, these results are of especially wide relevance.

Another contribution of this thesis is to add an important piece of geographical information to the mapping of population ethics’ dangerous waters. Virtually all theories of population axiology have focused on avoiding the Repugnant Conclusion, according to which, for any population, there is a better population consisting of lives that are barely good. In Chapter 8, I urge theorists to focus on another disturbing implication for theories of population axiology: the Monstrous Conclusion, according to which, for any population, there is a better population consisting of a single individual. In the Chapter I demonstrate what the necessary intellectual cost is to avoid this Monster, namely the Absolute Priority Condition. I then defended that Asymptotic Prioritarianism is a plausible way to avoid the Monstrous Conclusion. Still, I suspect it is safer to sail away from the Scylla represented by the Absolute Priority Condition than from the Charybdis represented by the Monstrous Conclusion.

⁹⁵ While Parfit never referred to the Pillars of Hercules, he too compared his researches to an “open sea”. Both the opening of (Parfit, 1986) and the last page of (Parfit, 2017b) are dominated by the following quote from § 343 of the fifth book of Nietzsche’s *The Gay Science*: “At last the horizon appears free to us again, even granted that it is not bright; at last our ships may venture out again, venture out to face any danger; all the daring of the lover of knowledge is permitted again; the sea, our sea, lies open again; perhaps there has never been such an ‘open sea’” (Nietzsche, 2006).

This thesis, however, offers more than just negative results. The thesis presents two key, positive findings as to how to navigate the dilemmas of population ethics. The first is that, in Chapter 6: Barely good lives and the Intuition of Neutrality, I show that the most prominent traditions for identifying good, bad and neutral lives imply what is known as the Intuition of Neutrality, according to which there is a range of wellbeing levels where it is not better nor worse to live or to create a life at these levels. Thus, I suggest that the safest way to navigate the dangerous waters of population ethics is a theory that respects the Intuition of Neutrality.

The major positive finding of this thesis is the Structured Range View, defended in Chapter 7: The Structured Range View. This theory joins a “totalist” ordering for ranking populations with a model of the Intuition of Neutrality. According to the “totalist” ordering, a population is better than another if it contains more total welfare. According to the model of the Intuition of Neutrality, for a range of welfare levels, it is permitted, but not required, to create lives at these levels, provided that they do not fall short of value with respect to the alternatives. I show that this theory respects most adequacy conditions for any theory of population ethics. I argued that the adequacy condition that is violated by the Structured Range View, namely Non-Anti-Egalitarianism, can be permissibly, and possibly even desirably, violated precisely in the way the Structured Range View does it. I further showed that the Structured Range View, originally formulated to aggregate population value, can be extended to aggregate value within a life and be equally plausible. I ultimately argue that we should adopt the Structured Range View as our theory for population axiology, and as our theory to aggregate value within a life.

While I have described how the Structured Range View provides many axiological answers, its practical implications are yet to be explored. Crucially, we need research as to which lives are in the neutral range, which are below, and which are above.

Once the lives within and outside of the neutral range have been identified, important applications of the Structured Range View are to be expected in the morality towards non-human animals and climate responsibilities. For example, some plausible candidates for lives living at a

neutral welfare level, if they have any welfare at all, are insects and plants. One possible implication of the Structured Ranged Theory may be that the arguments on animal welfare against farming animals, such as cows and sheep, for human consumption, still allow for farming sufficiently simple insects. Another possible implication is that, while we may have some responsibility to save plants and plant species from the threat of climate change (Stroppa, 2023), this responsibility can never trump the responsibilities we have towards our fellow humans.

We can expect the Structured Range View to have implications for humans, too, particularly when it comes to bioethics. The version of the Structured Range View that aggregates population value is likely to give insights as to how to triage between different treatments, where treatments aimed at improving wellbeing levels outside the “neutral range” should be prioritized over treatments aimed at improving wellbeing levels within the “neutral range”. Additionally, the version of the Structured Range View that aggregates value within a single life seems to have a great number of implications. For example, one implication is that, for some patients whose lives will be at a neutral wellbeing level until the end of their lives, there are multiple, equally rationally permissible treatments. Some permissible treatments available to a patient allow the patient to live for longer, but at a lower (neutral) wellbeing level. Other permissible treatments allow the patient to live at a higher (neutral) wellbeing level, but for a shorter time. In general, the presence of a *range* of wellbeing levels at which it is neither better nor worse to live than not to live, is likely to have important implications in beginning-of-life and end-life debates in bioethics.

The research focusing on which welfare levels are in the neutral range, which are below, and which are above is assured to be difficult. However, I expect it to be only one of the difficulties when working out the practical implications of the Structured Range view. Another important issue to understand when looking for the practical applications of any axiological theory is how to deal with uncertainty and risk aversion. There is some recent exploration on how to join population axiology and decision theory under uncertainty or risk aversion, but it has thus far lead to further impossibility theorems and disturbing implications (Arrhenius & Stefánsson, 2023; Beckstead & Thomas, 2023;

Pettigrew, forthcoming; Stefánsson, Forthcoming; Thornley, 2021). Exploring the applications of population axiology seems as challenging as exploring population axiology itself.

And yet, we can be confident that theorists will nonetheless sail for that difficult exploration. Our responsibility towards future generations calls for a deeper understanding of population axiology and its applications. In addition to the need for moral guidance, the relentless pursuit of knowledge that is so deeply rooted in human nature is likely to spur further research in population axiology.

Indeed, according to Dante, responsibility towards future generations and pursuit of knowledge were the reasons that pushed Odysseus and his crew to venture beyond the Pillars of Hercules. Dante reports that, to convince his crew to leave the comfort of their homes to explore a sea of monsters, Odysseus uses the following words, that are as appropriate for sailors as they are for population ethicists:

Considerate la vostra semenza:	Consider ye thine lineage and offspring:
fatti non foste a viver come bruti,	Ye were not made to live like unto brutes,
ma per seguir virtute e canoscenza.	But for pursuit of virtue and of knowledge. ⁹⁶

⁹⁶ Most of the English translation reported here is by Longfellow (Dante Alighieri, 2017, vv. 118–121, canto XXVI). However, I modified the first verse of Longfellow’s translation, as Longfellow chose a more literal translation rather than a more metaphorical one. Longfellow’s translation says: “Consider ye the seed from which ye sprang”. This interprets Dante’s original term “semenza” in an excessively restrictive manner. While the literal translation of the term “semenza” is “seeds”, it metaphorically means “lineage”. Longfellow’s translation is obviously trying to preserve the metaphorical meaning, but manages to do so only partially. In fact, “semenza” means both one’s past lineage (one’s parents, grandparents, grand-grandparents...) and one’s future lineage (one’s children, grandchildren, grand-grandchildren...). In Longfellow’s translation, it is hard to read reference to one’s future lineage, which is what population ethicists are typically most concerned with. I use the wording “lineage and offspring” to further stress that Odysseus is asking to consider both past and future lineage (and to better respect the iambic pentameter). Of course, “lineage” and “offspring” are synonyms, which makes my verse translation less moving than Longfellow’s: there seems to be no way not to butcher a translation of Dante’s original. Maybe there is an impossibility theorem to be written about this, too.

(Dante Alighieri, 2019, pp. 118–121, Canto XXVI)

Appendix 1. On some attempts to distinguish Drab and Barely Conscious Lives from Very Short Lives and Rollercoaster Lives

Two putative explanations for the preference of Rollercoaster Lives and Very Short Lives over Drab Lives and Barely Conscious Lives have received attention.⁹⁷ One has to do with the view called Perfectionism, according to which “Even if some change brings a great net benefit to those who are affected, it is a change for the worse if it involves the loss of one of the best things in life” (Parfit, 2004, p. 26), that is, what we will call “Perfectionist Goods”. On Perfectionism, we should consider Rollercoaster Lives and Very Short Lives, that contain the best things in life, better than Drab Lives and Barely Conscious Lives, that do not. However, I show that this view implies that it can be better to have a *bad* life with respect to wellbeing if it contains a minimal surplus of the best things in life over the worst things in life (Section 1).

Another explanation that has been proposed as to why Rollercoaster Lives and Very Short Lives may be better than Drab Lives and Barely Conscious Lives is the presence of imprecision in the betterness ordering. For example, some have argued that the relation “better than” is multidimensional (Rabinowicz 2022b, Thornley 2022), or that some lives have undistinguished value (Gustafsson 2020). This may justify that lives achieving higher welfare levels at a time, such as Rollercoaster Lives and Very Short Lives, may be a preferable Z-population than lives that do not, such as Drab Lives and Barely Conscious Lives, for example because multidimensionality creates incommensurability between A and Z lives if Z-lives are Rollercoaster Lives and Very Short Lives, while any amount of Drab Lives and Barely Conscious Lives should be worse than A-lives. I find this solution unsatisfying too as I show that, for sufficiently simple lives, this theory implies that enough of these simpler lives are better than any number of more complex lives (Section 2).

⁹⁷ Some may suspect that what makes Rollercoaster Lives stick out is that an increased complexity in itself increases the value of a life. This would indeed make Rollercoaster Lives better than Drab Lives. However, if this were the case, we can re-establish equality between Drab and Rollercoaster Lives by adding further suffering to Rollercoaster Lives. So we must look for an explanation elsewhere.

1. On Perfectionist goods

Parfit calls Perfectionism the view according to which “Even if some change brings a great net benefit to those who are affected, it is a change for the worse if it involves the loss of one of the best things in life” (Parfit, 2004, p. 26). For example, if a change from a population to another were to bring a net benefit, but “Mozart's music would be lost”, or “Venice would be destroyed”, this would be a change for the worse overall (Parfit, 2004, p. 26). The core of perfectionism is that “what is good, ultimately, is the development of human nature” (Hurka, 1996, p. 3): while making people’s life may be part of human nature, so is creation and appreciation of artistic masterpieces and relationships. Thus, the presence of some “perfectionist good”, such as an artistic achievement or a relationship, can justify that some people’s wellbeing is not increased.

Perfectionism has a strong, long tradition in moral, political philosophy and wellbeing theory, at least according to contemporary Perfectionists like Hurka, who writes that “Perfectionism is the morality of Plato, Aristotle, Thomas Aquinas, G. W. Leibniz, G. W. F. Hegel, Karl Marx, Friedrich Nietzsche, F. H. Bradley, and (in part) G. E. Moore and Hastings Rashdall” (Hurka, 1987, p. 727). The first to suggest that Perfectionism may avoid the Repugnant Conclusion is James Griffin, who writes that

Perhaps it is better to have a certain number of people at a certain high level than a very much larger number at a level where life is just worth living. Then we might wish to stop the slide [from A-populations towards larger Z-populations] . . . at that point along the line where people’s capacity to appreciate beauty, to form deep loving relationships, to accomplish something with their lives beyond just staying alive [...] all disappear. (Griffin, 2002, p. 340)

Griffin is assuming two things. First, that people in the A-population are enjoying such Perfectionist Goods. This is controversial: A-people may be experiencing immense pleasure, for example because

plugged into an experience machine. But let us bracket this issue and simply assume that Griffin's argument is restricted for A-populations where there are such high goods.

Second, Griffin assumes that people in the Z-Population are not enjoying these goods. But people who live Rollercoaster Lives in the Z-Population may indeed be enjoying such goods. So Griffin's argument to avoid the Repugnant Conclusion does not work against reaching a Repugnant Conclusion with Rollercoaster Lives. Clearly, it does not work against reaching a Repugnant Conclusion with Very Short Lives either, as these lives may be filled by appreciation of beauty or meaningful relationships, despite it is hard to imagine how. Instead, it works really well against reaching a Repugnant Conclusion with Drab Lives or Barely Conscious Lives, where these goods are excluded. This seems to be exactly what we are looking for in this Section.

Applied to population ethics, Perfectionism is supposed to work as follows, in a nutshell:

The best things in life [or Perfectionist Goods] have more than just welfare value, and a sufficient quantity of these things give a life containing them an, all things considered, value that is lexically superior to that of any amount of welfare on its own. (Beard 2020, Section 2).

Parfit is aware that Perfectionism has two problems. First, as all goods, Perfectionist Goods come in a smooth spectrum.

If we merely compare Mozart and muzak, these two may also seem to be in quite different categories. But there is a fairly smooth continuum between these two. Though Haydn is not as good as Mozart, he is very good. And there is other music which is not far below Haydn's, other music not far below this, and so on. Similar claims apply to the other best experiences, activities, and personal relationships, and to the other things which give most to the value of life. Most of these things are on fairly smooth continua, ranging from the best to the least good. Since this is so, it may be hard to defend the view that what is best has more value—or

does more to make the outcome better—than any amount of what is nearly as good. (Parfit 2016, p. 130)

In other words, it seems particularly hard to justify that, if we could greatly improve people’s lives by listening to Haydn rather than the slightly better Mozart, we should not do it.

A committed Perfectionist may want to avoid this problem by arguing that what Parfit is worrying about is just a Sorites paradox. The Perfectionist may argue that, somewhere, there is a threshold separating Perfectionist Goods from non-Perfectionist Goods: we can lose Mozart for Haydn in order to improve people’s lives, but somewhere there will be a pair of items along the continuum where we cannot justify going from the better item to the worse item, not even if this were to improve people’s life. This threshold is hard to spot, the Perfectionist may argue, because we are dealing with a Sorites paradox. Let us grant it works for now.

The second problem is that Perfectionism struggles to accommodate “the moral importance of relieving or preventing great suffering. We should reject the Nietzschean view that the prevention of great suffering can be ranked wholly below the preservation of creation of the best things in life.” (Parfit 2004, p. 163).

To understand this problem, consider the following. Suppose that we were in the position of creating a building that would be an artistic achievement as Perfect, if not better, than the Pyramids in pristine condition. However, in order to do so, we would need to deploy a mass of slaves larger than the one used to build the Pyramids. The amount of effort required by each slave will take all the time in their life, preventing them from doing anything but work and sleep. The physical exhaustion of each of these slaves is sure to kill them with a horrible death, and each slave riot will need to be repressed in inhumane ways to ensure the completion of the building.

As Parfit recognizes, this result is morally unacceptable. (Parfit 2004) wants to bracket the problem, insisting that “since we can assume that in the various outcomes we are considering there would be no such suffering” (p. 164), but any candidate theory of value aggregation wants to be able

to accommodate cases involving suffering. In the Rollercoaster version of the Z-population, there *is* enormous suffering together with enormous goods: a theory for value aggregation wants to be able to suggest whether this is better or worse than A. So Perfectionists need to say something about whether and when Perfectionist Goods can justify suffering.

There are two main answers to that problem. They both have implausible implications.

We call the first of them *Beard's solution* as it has been explicitly defended by Simon Beard (Beard, 2020). On this view, there are some “Perfectionist Bads” such as slavery and great suffering, that can balance Perfectionist Goods. A Perfectionist Bad mirrors a Perfectionist Good: it is something that has more than just welfare value, and a sufficient quantity of these Bads give a life containing them an, all things considered, value that is lexically inferior to that of any amount of negative welfare on its own. These “Perfectionist Bads” need not to be the opposite of the “development of human nature” as Hurka intends it (Hurka, 1996, p. 3): they may be simply serious enough bads.

On this view, for any amount of Perfectionist Goods, there is some amount of Perfectionist Bads such that it is worse to have both these many Perfectionist Goods and these many Perfectionist Bads than to have none of them. This mitigates the Nietzschean conclusion that we should pursue great art and relationships even at the cost of great personal suffering: in fact, this pursue would be justified only if the Perfectionist Goods are sufficiently weightier than the Perfectionist Bads.

This solution, however, has bizarre implications. Since Perfectionists want Perfectionist Goods to be lexically superior to “normal goods”, they believe that any surplus of Perfectionist Goods is better than any surplus of normal goods. But, if this is true, then it can be good to bring people into existence whose lives would be *bad*, all things considered, provided that they have the slightest surplus of Perfectionist Good over Perfectionist Bad. For example, fill a person’s life with any bad that is not perfect, such as boredom, headache and slight nausea, and give them one Perfectionist Good, say seeing the Pyramids once, and a Perfectionist Bad, say being a slave for one day in building the Pyramids, such that the Perfectionist Good is slightly better than the Perfectionist Bad is bad

(perhaps the sight of the pyramids is during a particularly beautiful sunset). On this Perfectionist view, for any population, there is *better* a population of such boredom-and-nausea-filled lives with a slight surplus of Perfectionist Good over Perfectionist Bad. This is unacceptable.⁹⁸

The other solution is to bite the bullet and accept the *Nietzschean Solution*. There are no Perfectionist Bads, and the obtainment of the best things in life can justify sufferings of any extent – even inflicting the greatest atrocities on people. I expect most people, like Parfit, to balk at this view just for this fact alone. But I want to offer an additional challenge to it, to fully capture how implausible this view is.

Recall that Perfectionist Goods are on a spectrum. There is a smooth continuum that goes from any Perfectionist Good to just a normal good – one can go from Mozart’s music to muzak in arbitrarily small steps. Now, the Perfectionist has to say that there will be a pair of items in the smooth continuum, one just slightly better than the other, such that one is a Perfectionist Good and the next one is not. Maybe the next one is a “normal” good, maybe it is metaphysically vague (i.e. there is no fact of the matter as to) whether this next item is a Perfectionist Good or not. This means that there is a pair of items in the spectrum, differing only slightly, such that obtaining one definitely justifies any amount of suffering, and the other either does not or there is no fact of the matter as to whether it does. For example, there are two records of some obscure Haydn Symphony, one slightly better recorded than the other (eg. the microphone was one centimeter too close to the strings), such that producing one record justifies mass slavery and torture, while producing the other one does not. This is deeply implausible (and an instance of hypersensitivity: see Chapter 3. The Benign Addition Paradox).

We conclude that Perfectionism cannot explain why a Repugnant Conclusion with a Z-Population consisting of Rollercoaster Lives is more plausible than a Repugnant Conclusion with a Z-Population consisting of muzak and Potatoes lives, at least not without unacceptable consequences.

⁹⁸ And is a version of the Sadistic Conclusion: Chapter 3, Section 3.

We need to look for something else. However, before proposing my own view, I need to examine another candidate theory that has been recently proposed in the literature: the idea that there is imprecision in the betterness ordering.

2. Imprecision in the betterness ordering

It is plausible that the betterness ranking is not precise. If this is true, then the wellbeing level of Rollercoaster Lives is just not the same of the wellbeing level of Drab Lives: this opens the possibility that we may have a Repugnant Conclusion where everyone in population Z is living Rollercoaster Lives, but not a Repugnant Conclusion where everyone in population Z with Drab Lives. This idea has been recently explored, in similar ways but with different justifications, in (Gustafsson, 2020; Rabinowicz, 2022b; Thornley, 2022). In this subsection, I use Thornley's theory as a token theory to illustrate how, as all these authors admit, while each of them succeeds in stating that the wellbeing level of Rollercoaster Lives is just not the same of the wellbeing level of Drab Lives, they all fail in avoiding a Repugnant Conclusion where everyone in population Z lives Drab Lives.

All the theories I explore in this Appendix are Critical Range Theories. This is a family of theories introduced in Chapter 5: Totalism: Between Repugnance and Greediness. According to this Critical Range Theories, there is a range of wellbeing levels that is not better nor worse than nonexistence. While these theories are structurally identical, they are motivated in different ways. In all cases, they are redescribing barely good lives, distancing themselves from classical "totalism", by adding some "imprecision" to the ranking of lives. In all cases, they manage to justify some incommensurability, but admittedly get the same problem: this imprecision leads to a Repugnant Conclusion composed of sufficiently simple beings. Let us now explore Thornley's theory as an example of these theories.

The idea that there is imprecision in the betterness ranking is intuitively compelling. As Thornley puts it, while some trade-offs are worth making, such as "going to the dentist to prevent

tooth decay” (Thornley 2022, pg. 20), for some others there seems to be no fact of the matter as to whether or not they are worth making. For example,

A parent says to their child, “No dessert unless you finish your dinner.” The child knows exactly what finishing dinner involves. They are all too familiar with the taste of peas and can see one hundred of them left on the plate. They also know what dessert will be like. The jelly is sitting on the counter and promises to taste as good as it always has. In this case, the trade-off may be neither worth making nor worth not making. And a small improvement to the child’s predicament need not resolve the issue. Suppose that the parent takes pity on the child and removes one pea from the plate. That need not ensure that finishing dinner is now a trade-off worth making. (Thornley 2022, pg. 20-21)

This supports the intuition that “certain goods [are] incommensurable with other goods, certain bads incommensurable with other bads, and certain combinations of goods and bads incommensurable with other combinations” (Thornley 2022, pg. 21), where “incommensurable” means that there is no fact of the matter as to whether the items are equally good, or one better than the other. If certain goods are incommensurable with certain others, then two lives that significantly differ with respect to some goods may be incommensurable.

However, of course, not all lives are incommensurable with one another: if two lives are perfectly identical but for the presence of a stubbed toe, the life without the stubbed toe is certainly better. Incommensurability does not prevent ranking lives, and in fact, Thornley proposes a ranking as follows.

He suggests there may be many dimensions of value along which we can compare lives: pleasure, pain, achievement, relationships and so on. Goods may be incommensurable because they improve a life with respect to different dimensions of value. To rank lives, we need to see how they compare in all dimensions. Specifically, he suggests ranking lives as follows:

- A life L1 is better than another life L2 iff L1 is better than L2 in at least one dimension and worse in none.
- A life L1 is worse than another life L2 iff L1 is worse than L2 in at least one dimension and better in none.
- Lives L1 and L2 are equally good iff they are equally good along all dimensions.
- Lives L1 and L2 are incommensurable iff there is at least one dimension in which L1 is better than L2, and at least one dimension in which L1 is worse than L2.

This way, Drab Lives may be incommensurable with Rollercoaster Lives. In fact, Drab Lives may be worse than Rollercoaster Lives with respect to one dimension (say, achievement), but better with respect to another (say, pain): the two lives are not at the same wellbeing level, but none is better than the other. This allows for the possibility of treating Drab Lives and Rollercoaster Lives differently when it comes to the Repugnant Conclusion.

However, Thornley's model does not quite do that. Consider a very simple being, say, an insect, whose experience of value dimensions is very limited to one dimension only: suppose it can only feel pleasure. Thornley's theory implies a Repugnant Conclusion where, for any population, there is a better population consisting of sufficiently many of these insects, that simply experience a mild pleasure for the entire duration of their existence (Thornley 2022, pg. 30/31): these are Drab Lives. Thus, Thornley's "imprecise exchange rates" view does not manage to avoid a Repugnant Conclusion of Drab Lives. All other "imprecise" theories I mentioned, despite having other theoretical virtues, admit they face the same challenge as Thornley, and fail at that, as admitted by (Rabinowicz, 2022b, pp. 130–136).

Appendix 2. List of conditions and conclusions

Absolute Priority Condition: Let w_i and w_j be two welfare levels, Δw be an increase in welfare, and m a multiplier. $\forall w_i, \forall \Delta w$ and $\exists w_j > w_i$ s. t. $\forall m > 1, g(w_i + \Delta w) - g(w_i) > g(w_j + m\Delta w) - g(w_j)$

Absolute Priority Condition (informal): for any amount of moral importance generated by giving some welfare increase Δw to some individual at welfare level w_i , there is some vastly better off individual at welfare level w_j such that, no matter how big the multiplier m of the welfare increase Δw is, it is more important to give Δw to w_i , than $m\Delta w$ to w_j .

Benign Addition. For any two populations P_1 and P_2 , if the wellbeing level of everyone in P_1 is higher in P_2 , and there are additional people with a non-negative wellbeing level, then P_2 is better.

Cyclical incommensurability. Let option B be not better, not worse, and not as good as option A, and let B be much better than option C. This is insufficient to conclude that C is worse than A.

Greediness: “The net effect of one bad thing and one neutral thing should be bad” (Broome 2004, p. 170)

Irreducibility. Some lives are better than any life reduced to containing only one value, such as pleasure.

Maximal Repugnance: Let life x and life y be lives that are identical, except that y has one fewer gumdrop’s worth of pleasure and one more hangnail’s worth of pain than x , such that

(1) each population of wonderful lives is worse than some population of x lives and (2) each population of awful lives is better than some population of y lives. (Adapted from (Thornley, 2022, p.11))

Maximal Greediness: for any population of awful lives and any population of wonderful lives, (1) there is some population of straightforwardly-better-than-blank lives such that the population of awful lives is not worse than the population of wonderful lives plus the straightforwardly- better-than-blank lives, or (2) there is some population of straightforwardly- worse-than-blank lives such that the population of wonderful lives is not better than the population of awful lives plus the straightforwardly-worse-than-blank lives. (Thornley 2022, p. 17).

Monstrous Conclusion: for any population, there is a better population consisting of just one individual.

Non-Anti-Egalitarianism. For any two populations P_1 and P_2 , if P_2 has higher average wellbeing than P_1 , higher total wellbeing than P_2 , and perfect equality, then P_2 is better than P_1 .

Non-Sadism. It cannot be better to add bad lives than good lives to a population.

Non-Triviality. For some fixed slight decrease d in one value, there is some sufficiently great increase I in another value such that, whatever a life is like, it is at least as good (that is, better, or equally good) to slightly decrease one of its values by d to greatly increase another value by I .

Pain Non-Triviality. For some fixed slight increase d in one value, there is some sufficiently great increase I in pain such that, whatever a life is like, it is at least as bad (that is, worse, or equally good) to slightly increase one of its values by d to greatly increase the amount of pain in one's life by I .

Quality Condition: There is a perfectly equal population with very high positive welfare which is at least as good as any population with very low positive welfare. (Arrhenius, 2000b, p. 41)

Repugnant Conclusion: For each population A of very many people— say, ten billion—all of whom have a very high quality of life, there is a better population Z consisting of some much larger number of people who would have lives that are barely good. (Adapted from Parfit, 2004, p.10)

Restriction to a range: if option X is incommensurable with option Y, for any option P much better than Y, P is better than X, and for any option Q much worse than Y, Q is worse than X.

Separability of lives. The goodness of conferring some benefit on one person, or of bringing some people into existence, should not depend on how many other people enjoy that benefit or already exist—e.g., on distant planets.

Trade-off Condition: Let Δw be any welfare difference, let w_i and w_j be any two wellbeing levels, and let k be any multiplier. $\forall k > 1 \forall \Delta w$ and $\forall w_i, \exists w_j > w_i$ s. t. $g(w_i + \Delta w) - g(w_i) > g(w_j + k\Delta w) - g(w_j)$

Trade-off Condition (Informal): for any multiplier k , any welfare increase Δw and any welfare level w_i , there is some sufficiently better off welfare level w_j such that giving the welfare increase Δw to the person at the given welfare level w_i is better than giving the multiplied

welfare increase to the better off individual (even though the multiplier may be arbitrarily large).

Transitivity of "better than": If P_1 is better than P_2 and P_2 is better than P_3 , then P_1 is better than P_3 .

Weak Repugnant Conclusion: For any population, there is a better population consisting of people just above the critical range.

Weak Reverse Repugnant Conclusion: For any population, there is a worse population consisting of people just below the critical range.

Bibliography

- Adler, M. D. (2009). Future Generations: A Prioritarian View. *George Washington Law Review*, 77, 1478–1520.
- Adler, M. D. (2013) The Pigou-Dalton Principle and the Structure of Distributive Justice. Working paper at Duke University School of Law (May 1, 2013). Available at SSRN: <https://ssrn.com/abstract=2263536> or <http://dx.doi.org/10.2139/ssrn.2263536>
- Adler, M. D. (2019). *Measuring Social Welfare: An Introduction* (1st ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780190643027.001.0001>
- Adler, M. D., & Holtug, N. (2019). Prioritarianism: A response to critics. *Politics, Philosophy & Economics*, 18(2), 101–144. <https://doi.org/10.1177/1470594X19828022>
- Adler, M. D., & Norheim, O. F. (Eds.). (2022). *Prioritarianism in practice*. Cambridge University Press.
- Ainslie, G. (2001). *Breakdown of will*. Cambridge University Press.
- Algander, P. (2012). A Defence of the Asymmetry in Population Ethics. *Res Publica*, 18(2), 145–157. <https://doi.org/10.1007/s11158-011-9164-0>
- Andersson, H. (2015). Propping Up the Collapsing Principle. *Ethical Theory and Moral Practice*, 18(3), 475–486. <https://doi.org/10.1007/s10677-014-9533-9>
- Andersson, H., & Herlitz, A. (2018). A More Plausible Collapsing Principle. *Theoria*, 84(4), 325–336. <https://doi.org/10.1111/theo.12166>
- Andersson, H., & Herlitz, A. (Eds.). (2022). *Value incommensurability: Ethics, risk. and decision-making*. Routledge.
- Andreou, C. (2006). Environmental Damage and the Puzzle of the Self-Torturer. *Philosophy & Public Affairs*, 34(1), 95–108. <https://doi.org/10.1111/j.1088-4963.2006.00054.x>
- Andreou, C. (2007). Understanding Procrastination. *Journal for the Theory of Social Behaviour*,

- 37(2), 183–193. <https://doi.org/10.1111/j.1468-5914.2007.00331.x>
- Arneson, R. J. (2000). Perfectionism and Politics. *Ethics*, 111(1), 37–63.
<https://doi.org/10.1086/233418>
- Arneson, R. J. (2002). Why Justice Requires Transfers to Offset Income and Wealth Inequalities. *Social Philosophy and Policy*, 19(1), 172–200. <https://doi.org/10.1017/S0265052502191084>
- Arrhenius, G. (2000a). An Impossibility Theorem for Welfarist Axiologies. *Economics and Philosophy*, 16(2), 247–266. <https://doi.org/10.1017/S0266267100000249>
- Arrhenius, G. (2000b). *Future Generations: A challenge for moral theories*. Uppsala University Press.
- Arrhenius, G. (2005). Superiority in Value. *Philosophical Studies*, 123(1–2), 97–114.
<https://doi.org/10.1007/s11098-004-5223-0>
- Arrhenius, G. (2016). Population Ethics and Different-Number-Based Imprecision. *Theoria*, 82(2), 166–181. <https://doi.org/10.1111/theo.12094>
- Arrhenius, G. (2022). Population Paradoxes without Transitivity. In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (pp. 180–203). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190907686.013.11>
- Arrhenius, G. (Forthcoming). *Population Ethics: The challenge of future generations*. Oxford University Press.
- Arrhenius, G., & Andersson, E. (2022). The Repugnant Conclusion: An Overview. In S. M. Gardiner (Ed.), *The Oxford Handbook of Intergenerational Ethics* (1st ed.). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190881931.013.7>
- Arrhenius, G., Budolfson, M., & Spears, D. (2021). Does Climate Change Policy Depend Importantly on Population Ethics?: Deflationary Responses to the Challenges of Population Ethics for Public Policy. In G. Arrhenius, M. Budolfson, & D. Spears, *Philosophy and Climate Change* (pp. 111–136). Oxford University Press.
<https://doi.org/10.1093/oso/9780198796282.003.0006>

- Arrhenius, G., & Rabinowicz, W. (2005). Millian Superiorities. *Utilitas*, 17(2), 127–146.
<https://doi.org/10.1017/S0953820805001494>
- Arrhenius, G., & Rabinowicz, W. (2010). Better to be than not to be? In H. Joas & B. Klein, *The benefit of broad horizons: Intellectual and institutional preconditions for a global social science festschrift for Björn Wittrock on the occasion of his 65th birthday*. Brill.
- Arrhenius, G., & Rabinowicz, W. (2015a). *The Value of Existence* (I. Hirose & J. Olson, Eds.). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199959303.013.0023>
- Arrhenius, G., & Rabinowicz, W. (2015b). *Value Superiority* (I. Hirose & J. Olson, Eds.). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199959303.013.0013>
- Arrhenius, G., & Stefánsson, H. O. (2023). Population ethics under risk. *Social Choice and Welfare*.
<https://doi.org/10.1007/s00355-023-01452-8>
- Arrow, K. J. (1950). A Difficulty in the Concept of Social Welfare. *Journal of Political Economy*, 58(4), 328–346. <https://doi.org/10.1086/256963>
- Arrow, K. J. (1971). *Essays in the theory of risk-bearing*. Markham Pub. Co.
- Bader, R. M. (2022). The Asymmetry. In R. M. Bader, *Ethics and Existence* (pp. 15–37). Oxford University Press. <https://doi.org/10.1093/oso/9780192894250.003.0002>
- Baum, S. D. (2008). Better to exist: A reply to Benatar. *Journal of Medical Ethics*, 34(12), 875–876. <https://doi.org/10.1136/jme.2007.023986>
- Bayle, P. (1696). *Dictionnaire historique et critique* (1734th–1738th ed.). Pierre Des Maizeaux.
- Beard, S. (2020). Perfectionism and the Repugnant Conclusion. *The Journal of Value Inquiry*, 54(1), 119–140. <https://doi.org/10.1007/s10790-019-09687-4>
- Beard, S., & Kaczmarek, P. (2019). *On the wrongness of human extinction*.
<https://doi.org/10.14275/2465-2334/20199.BEA>
- Beckstead, N., & Thomas, T. (2023). A paradox for tiny probabilities and enormous values. *Nous*, nous.12462. <https://doi.org/10.1111/nous.12462>
- Benatar, D. (2008). *Better never to have been: The harm of coming into existence*. Clarendon press.

- Benatar, D. (2013). Still Better Never to Have Been: A Reply to (More of) My Critics. *The Journal of Ethics*, 17(1–2), 121–151. <https://doi.org/10.1007/s10892-012-9133-7>
- Benatar, D. (2017). *The human predicament: A candid guide to life's biggest questions*. Oxford University Press.
- Binmore, K., & Voorhoeve, A. (2003). Defending Transitivity against Zeno's Paradox. *Philosophy & Public Affairs*, 31(3), 272–279. <https://doi.org/10.1111/j.1088-4963.2003.00272.x>
- Blackorby, C., Bossert, W., & Donaldson, D. (1996). Quasi-orderings and population ethics. *Social Choice and Welfare*, 13(2), 129–150. JSTOR.
- Blackorby, C., Bossert, W., & Donaldson, D. (1997). Critical-Level Utilitarianism and the Population-Ethics Dilemma. *Economics and Philosophy*, 13(2), 197–230. <https://doi.org/10.1017/S026626710000448X>
- Blackorby, C., Bossert, W., & Donaldson, D. (2005). *Population issues in social-choice theory, welfare economics and ethics*. Cambridge University Press.
- Blackorby, C., Bossert, W., Donaldson, D., & Fleurbaey, M. (1998). Critical Levels and the (Reverse) Repugnant Conclusion. *Journal of Economics*, 67(1), 1–15. JSTOR.
- Boonin, D. (2020). *Non-identity problem and the ethics of future people*. OXFORD UNIV Press.
- Bossert, W., Cato, S., & Kamaga, K. (2022). Critical-level Sufficiency*^{*}. *Journal of Political Philosophy*, 30(4), 434–461. <https://doi.org/10.1111/jopp.12267>
- Bradford, G. (2015). *Achievement* (First edition). Oxford University Press.
- Bramble, B. (2016). A New Defense of Hedonism about Well-Being. *Ergo, an Open Access Journal of Philosophy*, 3(20201214). <https://doi.org/10.3998/ergo.12405314.0003.004>
- Bratman, M. E. (2012). Time, Rationality, and Self-Governance. *Philosophical Issues*, 22(1), 73–88. <https://doi.org/10.1111/j.1533-6077.2012.00219.x>
- Brennan, S. (2006). Moral Lumps. *Ethical Theory and Moral Practice*, 9(3), 249–263.
- Briggs, R., & Nolan, D. (2015). Utility Monsters for the Fission Age. *Pacific Philosophical Quarterly*, 96(3), 392–407. <https://doi.org/10.1111/papq.12079>

- Broome, J. (1991). *Weighing goods: Equality, uncertainty, and time*. Basil Blackwell.
- Broome, J. (1992). Deontology and Economics. *Economics and Philosophy*, 8(2), 269–282.
<https://doi.org/10.1017/S0266267100003072>
- Broome, J. (1997). Is incommensurability vagueness? In R. Chang (Ed.), *Incommensurability, incomparability, and practical reason* (pp. 131–149). Harvard University Press.
- Broome, J. (2003). *Ethics out of economics* (Transferred to digital print). Cambridge Univ. Press.
- Broome, J. (2004). *Weighing Lives* (1st ed.). Oxford University Press Oxford.
<https://doi.org/10.1093/019924376X.001.0001>
- Broome, J. (2005). Should We Value Population?*. *Journal of Political Philosophy*, 13(4), 399–413. <https://doi.org/10.1111/j.1467-9760.2005.00230.x>
- Broome, J. (2009). Reply to Rabinowicz. *Philosophical Issues*, 19, 412–417. JSTOR.
- Broome, J. (2012). *Climate matters: Ethics in a warming world*. W. W. Norton.
- Broome, J. (2015). Climate change: Life and death. In J. Moss (Ed.), *Climate Change and Justice* (1st ed., pp. 184–200). Cambridge University Press.
<https://doi.org/10.1017/CBO9781316145340.011>
- Broome, J. (2022). Loosening the Betterness Ordering of Lives: A Response to Rabinowicz. In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (pp. 141–157). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780190907686.013.18>
- Broome, J. (2023). *No Continuum Argument for the Repugnant Conclusion*.
- Brown, C. (2005). Priority or Sufficiency ...or both? *Economics and Philosophy*, 21(2), 199–220.
<https://doi.org/10.1017/S0266267105000568>
- Brown, C. (2011). Better Never To Have Been Believed: Benatar On The Harm Of Existence. *Economics and Philosophy*, 27(1), 45–52. <https://doi.org/10.1017/S0266267110000465>
- Brown, C. (2017). Reply to Benatar. *Journal of Ethics and Social Philosophy*, 7(3), 1–2.
<https://doi.org/10.26556/jesp.v7i3.155>

- Brown, J. L. D. (2019). Additive Value and the Shape of a Life. *Ethics*, 130(1), 92–101.
<https://doi.org/10.1086/704345>
- Buchak, L. (2017). *Risk and rationality* (First published in paperback). Oxford University Press.
- Buchanan, A. E., & Brock, D. W. (Eds.). (2007). *From chance to choice: Genetics and justice* (8th printing). Cambridge Univ. Press.
- Budolfson, M., & Spears, D. (2021). Population Ethics and the Prospects for Fertility Policy as Climate Mitigation Policy. *The Journal of Development Studies*, 57(9), 1499–1510.
<https://doi.org/10.1080/00220388.2021.1915481>
- Bykvist, K. (2007). The Benefits of Coming into Existence. *Philosophical Studies*, 135(3), 335–362. <https://doi.org/10.1007/s11098-005-3982-x>
- Bykvist, K. (2016). *Preference-Based Views of Well-Being* (M. D. Adler & M. Fleurbaey, Eds.; Vol. 1). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199325818.013.10>
- Cafaro, P. (2022). Climate ethics and population policy: A review of recent philosophical work. *WIREs Climate Change*, 13(2), e748. <https://doi.org/10.1002/wcc.748>
- Campbell, T., & Mosquera, J. (2020). Transformative experience and the shark problem. *Philosophical Studies*, 177(11), 3549–3565. <https://doi.org/10.1007/s11098-019-01382-1>
- Camus, A. (2004). *The plague ; The fall ; Exile and the kingdom ; and selected essays* (D. Bellos, Trans.). Alfred A. Knopf.
- Camus, A., & O'Brien, J. (1991). *The myth of Sisyphus and other essays* (1st Vintage international ed). Vintage Books.
- Carlson, E. (1998). Mere Addition and Two Trilemmas of Population Ethics. *Economics and Philosophy*, 14(2), 283–306. <https://doi.org/10.1017/S0266267100003862>
- Carlson, E. (2005). 'Intransitivity without Zeno's paradox.' In Toni Rønnow-Rasmussen & Michael J. Zimmerman (eds.), *Recent Work on Intrinsic Value*. Springer. pp. 273--277
- Carlson, E. (2007). Higher Values and Non-Archimedean Additivity. *Theoria*, 73(1), 3–27.
<https://doi.org/10.1111/j.1755-2567.2007.tb01185.x>

- Carlson, E. (2010). Generalized extensive measurement for lexicographic orders. *Journal of Mathematical Psychology*, 54(4), 345–351. <https://doi.org/10.1016/j.jmp.2010.06.002>
- Carlson, E. (2013). Vagueness, Incomparability, and the Collapsing Principle. *Ethical Theory and Moral Practice*, 16(3), 449–463. <https://doi.org/10.1007/s10677-012-9352-9>
- Carlson, E. (2022). On Some Impossibility Theorems in Population Ethics. In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (pp. 203–225). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780190907686.013.14>
- Casal, P. (2007). Why Sufficiency Is Not Enough. *Ethics*, 117(2), 296–326.
<https://doi.org/10.1086/510692>
- Chang, R. (Ed.). (1997). *Incommensurability, incomparability, and practical reason*. Harvard University Press.
- Chang, R. (2002). The Possibility of Parity. *Ethics*, 112(4), 659–688.
<https://doi.org/10.1086/339673>
- Chang, R. (2015a). *Value Incomparability and Incommensurability* (I. Hirose & J. Olson, Eds.). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199959303.013.0012>
- Chang, R. (2015b). Value Pluralism. In J. Wright, *International Encyclopedia of the Social and Behavioral Sciences (Second Edition)* (pp. 21–26). Elsevier.
- Chang, R. (2016). Parity, Imprecise Comparability and the Repugnant Conclusion. *Theoria*, 82(2), 182–214. <https://doi.org/10.1111/theo.12096>
- Chappell, R. Y. (2021). *Parfit's ethics*. Cambridge university press.
- Cohen, D. (2020). An Actualist Explanation of the Procreation Asymmetry. *Utilitas*, 32(1), 70–89.
<https://doi.org/10.1017/S0953820819000293>
- Constantinescu, C. (2012). Value Incomparability and Indeterminacy. *Ethical Theory and Moral Practice*, 15(1), 57–70. <https://doi.org/10.1007/s10677-011-9269-8>
- Constantinescu, C. (2016). Vague Comparisons. *Ratio*, 29(4), 357–377.

<https://doi.org/10.1111/rati.12146>

Cowen, T. (1996). What Do We Learn from the Repugnant Conclusion? *Ethics*, 106(4), 754–775.

<https://doi.org/10.1086/233671>

Cowen, T., & High, J. (1988). Time, bounded utility, and the St. Petersburg paradox. *Theory and Decision*, 25(3), 219–223. <https://doi.org/10.1007/BF00133163>

Cowie, C. (2020). *The repugnant conclusion: A philosophical inquiry*. Routledge.

Crisp, R. (1988). *Ideal utilitarianism: Theory and practice* [University of Oxford]. .

<http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.253770>

Crisp, R. (1992). Utilitarianism and the Life of Virtue. *The Philosophical Quarterly*, 42(167), 139.

<https://doi.org/10.2307/2220212>

Crisp, R. (2003). Equality, Priority, and Compassion. *Ethics*, 113(4), 745–763.

<https://doi.org/10.1086/373954>

Crisp, R. (2006). *Reasons and the good*. Oxford University Press.

Cullity, G. (2018). Weighing Reasons. In D. Star (Ed.), *The Oxford handbook of reasons and normativity* (First edition). Oxford University Press.

Dancy, J. (2004). *Ethics Without Principles* (1st ed.). Oxford University Press Oxford.

<https://doi.org/10.1093/0199270023.001.0001>

Dante Alighieri. (2017). *The divine comedy* (H. W. Longfellow, Trans.). Dover Publications.

Dante Alighieri. (2019). *La Divina Commedia* (G. Inglese, Ed.). Carocci editore.

Dillenberger, D., & Vijay Krishna, R. (2014). Expected utility without bounds—A simple proof.

Journal of Mathematical Economics, 52, 143–147.

<https://doi.org/10.1016/j.jmateco.2013.12.011>

Dorsey, D. (2009). Headaches, Lives and Value. *Utilitas*, 21(1), 36–58.

<https://doi.org/10.1017/S095382080800335X>

Dorsey, D. (2014). Equality-tempered prioritarianism. *Politics, Philosophy & Economics*, 13(1), 45–61. <https://doi.org/10.1177/1470594X13483479>

- Dorsey, D. (2015). The Significance of a Life's Shape. *Ethics*, 125(2), 303–330.
<https://doi.org/10.1086/678373>
- Dorsey, D. (2021). *A theory of prudence*. Oxford University press.
- Doyal, L. (2007). Is human existence worth its consequent harm? *Journal of Medical Ethics*, 33(10), 573–576. <https://doi.org/10.1136/jme.2006.020156>
- Edwards, R. B. (1979). *Pleasures and pains: A theory of qualitative hedonism*. Cornell university press.
- Elson, L. (2017). Incommensurability as Vagueness: A Burden-Shifting Argument. *Theoria*, 83(4), 341–363. <https://doi.org/10.1111/theo.12129>
- Fehige, C. (1988). A Pareto Principle for Possible People. In C. Fehige & U. Wessels (Eds.), *Preferences* (pp. 509–543). De Gruyter.
- Feinberg, J. (1986). Wrongful Life and the Counterfactual Element in Harming. *Social Philosophy and Policy*, 4(1), 145–178. <https://doi.org/10.1017/S0265052500000467>
- Feldman, F. (1995). Justice, Desert, and the Repugnant Conclusion. *Utilitas*, 7(2), 189–206.
<https://doi.org/10.1017/S095382080000203X>
- Finneron-Burns, E. (2016). Contractualism and the Non-Identity Problem. *Ethical Theory and Moral Practice*, 19(5), 1151–1163. <https://doi.org/10.1007/s10677-016-9723-8>
- Fishburn, P. C. (1976). Unbounded Utility Functions in Expected Utility Theory. *The Quarterly Journal of Economics*, 90(1), 163. <https://doi.org/10.2307/1886093>
- Fletcher, G. (2013). A Fresh Start for the Objective-List Theory of Well-Being. *Utilitas*, 25(2), 206–220. <https://doi.org/10.1017/S0953820812000453>
- Frankfurt, H. (1987). Equality as a Moral Ideal. *Ethics*, 98(1), 21–43. JSTOR.
- Frick, J. (2017). On the survival of humanity. *Canadian Journal of Philosophy*, 47(2–3), 344–367.
<https://doi.org/10.1080/00455091.2017.1301764>
- Frick, J. (2022). Context-Dependent Betterness and the Mere Addition Paradox. In J. Frick, *Ethics and Existence* (pp. 232–263). Oxford University Press.

<https://doi.org/10.1093/oso/9780192894250.003.0010>

Friedman, A. (2009). Intransitive Ethics. *Journal of Moral Philosophy*, 6(3), 277–297.

<https://doi.org/10.1163/174552409X433391>

Fumagalli, R. (2018). Eliminating ‘life worth living’. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 175(3), 769–792. JSTOR.

Gehrlein, W. V. (2006). *Condorcet’s Paradox*. Springer-Verlag Berlin Heidelberg.

Gert, J. (2003). Requiring and Justifying: Two Dimensions of Normative Strength. *Erkenntnis*, 59(1), 5–36. <https://doi.org/10.1023/A:1023930108900>

Gert, J. (2007). Normative Strength and the Balance of Reasons. *The Philosophical Review*, 116(4), 533–562. JSTOR.

Gert, J. (2016). The Distinction between Justifying and Requiring: Nothing to Fear. In E. Lord & B. Maguire (Eds.), *Weighing Reasons* (pp. 157–172). Oxford University Press.

<https://doi.org/10.1093/acprof:oso/9780199315192.003.0008>

Gibbard, A. (1992) “Interpersonal comparisons: preference, good, and the intrinsic reward of a life.” In *Foundations of Social Choice Theory*, edited by Jon Elster and Anand Hylland. Cambridge: Cambridge University Press.

Glover, J. (1990). *Causing death and saving lives* (Repr). Penguin Books.

Goodrich, J. (2016). Mereological Dominance and the Logic of Better-Than. *Utilitas*, 28(4), 361–367. <https://doi.org/10.1017/S0953820815000503>

Greaves, H. (2017). Population axiology. *Philosophy Compass*, 12(11), e12442.

<https://doi.org/10.1111/phc3.12442>

Greaves, H., & Cusbert, J. (2022). Comparing Existence and Non-Existence. In H. Greaves & J. Cusbert, *Ethics and Existence* (pp. 61–92). Oxford University Press.

<https://doi.org/10.1093/oso/9780192894250.003.0004>

Griffin, J. (2002). *Well-being: Its meaning, measurement, and moral importance* (reprint). Clarendon Pr.

- Grill, K. (2017). Asymmetric population axiology: Deliberative neutrality delivered. *Philosophical Studies*, 174(1), 219–236. <https://doi.org/10.1007/s11098-016-0678-3>
- Grill, K. (2023). The Sum of Averages: An Egyptology-Proof Average View. *Utilitas*, 35(2), 103–118. <https://doi.org/10.1017/S0953820822000450>
- Gustafsson, J. E. (2018). Does the Collapsing Principle Rule Out Borderline Cases? *Utilitas*, 30(4), 483–492. <https://doi.org/10.1017/S0953820818000079>
- Gustafsson, J. E. (2020). Population axiology and the possibility of a fourth category of absolute value. *Economics and Philosophy*, 36(1), 81–110. <https://doi.org/10.1017/S0266267119000087>
- Gustafsson, J. E. (2022a). *Money-pump arguments*. Cambridge University Press.
- Gustafsson, J. E. (2022b). Our Intuitive Grasp of the Repugnant Conclusion. In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (pp. 370–389). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190907686.013.25>
- Gustafsson, J. E. (2023). The possibility of undistinguishedness. *Philosophical Studies*, 180(2), 609–613. <https://doi.org/10.1007/s11098-022-01889-0>
- Gustafsson, J. E., & Rabinowicz, W. (2020). A Simpler, More Compelling Money Pump with Foresight. *The Journal of Philosophy*, 117(10), 578–589. <https://doi.org/10.5840/jphil20201171037>
- Hájek, A., & Rabinowicz, W. (2022). Degrees of commensurability and the repugnant conclusion. *Noûs*, 56(4), 897–919. <https://doi.org/10.1111/nous.12388>
- Hájek, A., & Smithson, M. (2012). Rationality and indeterminate probabilities. *Synthese*, 187(1), 33–48. JSTOR.
- Handfield, T., & Rabinowicz, W. (2018). Incommensurability and vagueness in spectrum arguments: Options for saving transitivity of betterness. *Philosophical Studies*, 175(9), 2373–2387. <https://doi.org/10.1007/s11098-017-0963-9>

- Hare, C. (2010). Take the sugar. *Analysis*, 70(2), 237–247. JSTOR.
- Harney, J., & Khawaja, J. (2023). How to Assess Claims in Multiple-Option Choice-sets. *Philosophy & Public Affairs*, 51(1), 60–92. <https://doi.org/10.1111/papa.12227>
- Haybron, D. M. (2016). *Mental State Approaches to Well-Being* (M. D. Adler & M. Fleurbaey, Eds.; Vol. 1). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199325818.013.11>
- Hedden, B., & Muñoz, D. (2023). Dimensions of Value. *Noûs*, nous.12454.
<https://doi.org/10.1111/nous.12454>
- Heathwood, C. (2019). Which Desires Are Relevant to Well-Being? *Noûs*, 53(3), 664–688.
<https://doi.org/10.1111/nous.12232>
- Heathwood, C. (2020). An opinionated guide to “What Makes Someone’s Life Go Best”. In A. Sauchelli (Ed.), *Derek Parfit’s Reasons and Persons* (1st ed., pp. 94–114). Routledge.
<https://doi.org/10.4324/9780429488450-7>
- Hedberg, T. (2023). Climate Change and Population Ethics. In G. Pellegrino & M. Di Paola (Eds.), *Handbook of Philosophy of Climate Change* (pp. 1–16). Springer International Publishing.
https://doi.org/10.1007/978-3-030-16960-2_64-1
- Heyd, D. (1988). Procreation and value can ethics deal with futurity problems? *Philosophia*, 18(2–3), 151–170. <https://doi.org/10.1007/BF02380074>
- Heyd, D. (1994). *Genethics: Moral issues in the creation of people* (1. paperback print). Univ. of California Press.
- Hirose, I. (2016). Axiological Sufficiency. In C. Fourie & A. Rid (Eds.), *What is Enough?* (pp. 51–68). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199385263.003.0004>
- Holtug, N. (2010a). On the Value of Coming into Existence. In N. Holtug, *Persons, Interests, and Justice* (1st ed., pp. 129–150). Oxford University PressOxford.
<https://doi.org/10.1093/acprof:oso/9780199580170.003.0005>

- Holtug, N. (2010b). *Persons, interests, and justice*. Oxford University Press.
- Holtug, N. (2017). Prioritarianism. In N. Holtug, *Oxford Research Encyclopedia of Politics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190228637.013.232>
- Holtug, N. (2022). Prioritarianism and Population Ethics. In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (pp. 37–62). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190907686.013.12>
- Homerous. (1924). *The Iliad* (A. T. Murray, Trans.; Vol. 9). William Heinemann, Ltd.
- Homerus. (500). *Iliad. 1/12: Books I - XII* (Repr. 1999). Bristol Classical Press.
- Huemer, M. (2008). In Defence of Repugnance. *Mind*, 117(468), 899–933. JSTOR.
- Hume, D. (1990). *Dialogues concerning natural religion* (J. M. Bell, Ed.). Penguin Books.
- Hurka, T. (1987). The Well-Rounded Life. *The Journal of Philosophy*, 84(12), 727.
<https://doi.org/10.2307/2026589>
- Hurka, T. (1996). *Perfectionism* (1st ed.). Oxford University Press New York.
<https://doi.org/10.1093/0195101162.001.0001>
- Hurka, T. (2011). *Drawing Morals: Essays in Ethical Theory*. Oxford University Press.
<https://doi.org/10.1093/acprof:osobl/9780199743094.001.0001>
- Hurka, T. (2016). *Objective Goods* (M. D. Adler & M. Fleurbaey, Eds.; Vol. 1). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199325818.013.12>
- Jensen, K. K. (2008). Millian Superiorities and the Repugnant Conclusion. *Utilitas*, 20(3), 279–300.
<https://doi.org/10.1017/S0953820808003154>
- Kamm, F. (2015). Cost Effectiveness Analysis and Fairness. *Journal of Practical Ethics*, 3(1), 1–14.
- Kamm, F. M. (1985). Supererogation and Obligation: *Journal of Philosophy*, 82(3), 118–138.
<https://doi.org/10.2307/2026351>
- Kant, I. (1900). On the Failure of All Philosophical Trials in Theodicy. In *Immanuel Kant: Gesammelte Schriften (Akademie-Ausgabe)*. (Vol. 8, pp. 255–275). Georg Reimer, Walter de

Gruyter.

Kavka, G. S. (1982). The Paradox of Future Individuals. *Philosophy & Public Affairs*, 11(2), 93–112. JSTOR.

King James' bible. (1998). Book of Job. In *Old Testament*. Edinburgh : Canongate, 1998.
<https://search.library.wisc.edu/catalog/999887893402121>

King, W. (2022). *Essay of the Origin of Evil* (Classics reprint). FORGOTTEN BOOKS.

Kitcher, P. (2000). Parfit's Puzzle. *Noûs*, 34(4), 550–577. <https://doi.org/10.1111/0029-4624.00278>

Kosonen, P. (2022). *Tiny Probabilities of Vast Value* (PhD dissertation at Oxford University).

Kreps, D. M. (2013). *Microeconomic foundations. 1: Choice and competitive markets / David M. Kreps*. Princeton Univ. Pr.

Lee, A. Y. (2023). The Neutrality of Life. *Australasian Journal of Philosophy*, 101(3), 685–703.
<https://doi.org/10.1080/00048402.2022.2033284>

Leibniz, G. W. (2019). *Theodicy*. Anodos Books.

Lemos, N. M. (1993). Higher Goods and the Myth of Tithonus: *Journal of Philosophy*, 90(9), 482–496. <https://doi.org/10.2307/2940862>

Lenman, J. (2000). Consequentialism and Cluelessness. *Philosophy & Public Affairs*, 29(4), 342–370. <https://doi.org/10.1111/j.1088-4963.2000.00342.x>

Leopardi, G. (1992). *Dialogo di un venditore d'almanacchi e di un passeggero* (G. Celli, Ed.). Muzzio.

Levi, I. (1997). *The Covenant of Reason: Rationality and the Commitments of Thought* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139173032>

Lin, E. (2016). How to Use the Experience Machine. *Utilitas*, 28(3), 314–332.
<https://doi.org/10.1017/S0953820815000424>

Lin, E. (2022). Two Kinds of Desire Theory of Well-Being. *Midwest Studies in Philosophy*, 46, 55–86. <https://doi.org/10.5840/msp202212231>

Lin E. (forthcoming). 'Pleasure, Pain, and Pluralism about Wellbeing', *Philosophical Quarterly*.

- Magni, S. F. (2020). Person-affecting Procreative Beneficence. *Phenomenology and Mind*, 19, 124.
<https://doi.org/10.17454/pam-1909>
- Magni, S. F. (2021). In defence of person-affecting procreative beneficence. *Bioethics*, 35(5), 473–479. <https://doi.org/10.1111/bioe.12872>
- Maher, P. (1992). Diachronic Rationality. *Philosophy of Science*, 59(1), 120–141.
<https://doi.org/10.1086/289657>
- McCledden, E. F. (2012). *Rationality and dynamic choice: Foundational explorations*. Cambridge University Press : Lightning Source UK.
- McGee, V. (1999). An Airtight Dutch Book. *Analysis*, 59(4), 257–265. JSTOR.
- McMahan, J. (1981). Problems of Population Theory *Review of ‘Obligations to Future Generations’*. R. I. Sikora , Brian Barry. *Ethics*, 92(1), 96–127.
<https://doi.org/10.1086/292301>
- McTaggart, J. M. E., & Broad, C. D. (1968). *The nature of existence. Vol. 2* (Reprinted). At the University Press.
- Metz, T. (2002). Recent Work on the Meaning of Life. *Ethics*, 112(4), 781–814.
<https://doi.org/10.1086/340462>
- Mill, J. S. (2000). *Utilitarianism* (Broadview ed.). Broadview Press.
- Miller, D. (2021). Justice. In *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition). Edward N. Zalta. <https://plato.stanford.edu/archives/fall2021/entries/justice/>
- Mogensen, A. L. (2019). Staking our future: Deontic long-termism and the non-identity problem. *Global Priorities Institute Working Paper*.
- Mogensen, A. L. (2021). *Do not go gentle: Why the Asymmetry does not support anti-natalism* [GPI Working Paper No. 3-2021]. https://globalprioritiesinstitute.org/wp-content/uploads/Andreas-Mogensen_Do-not-go-gentle.pdf
- Mogensen, A. L. (2022). Against Large Number Scepticism. In A. L. Mogensen, *Ethics and Existence* (pp. 311–330). Oxford University Press.

<https://doi.org/10.1093/oso/9780192894250.003.0012>

Mori, M. (2013). *Manuale di bioetica: Verso una civiltà biomedica secolarizzata* (Nuova ed. ampliata). Le lettere.

Mori, M. (2014). *Introduzione alla bioetica 12 temi per capire e discutere* (Ristampa). Daniela Piazza.

Mulgan, T. (2001). What's Really Wrong with the Limited Quantity View? *Ratio*, 14(2), 153–164.

<https://doi.org/10.1111/1467-9329.00152>

Mulgan, T. (2002). The Reverse Repugnant Conclusion. *Utilitas*, 14(3), 360–364.

<https://doi.org/10.1017/S0953820800003654>

Muñoz, D. (2023). Sources of transitivity. *Economics and Philosophy*, 39(2), 285–306.

<https://doi.org/10.1017/S0266267122000074>

Nagel, T. (1971). The Absurd. *The Journal of Philosophy*, 68(20), 716.

<https://doi.org/10.2307/2024942>

Nagel, T. (2009). *Mortal questions* (14. print). Cambridge Univ. Press.

Narveson, J. (1967). Utilitarianism and New Generations. *Mind*, LXXVI(301), 62–72.

<https://doi.org/10.1093/mind/LXXVI.301.62>

Narveson, J. (1973). Moral Problems of Population: *Monist*, 57(1), 62–86.

<https://doi.org/10.5840/monist197357134>

Nebel, J. M. (2017). Priority, Not Equality, for Possible People. *Ethics*, 127(4), 896–911.

<https://doi.org/10.1086/691568>

Nebel, J. M. (2018). The Good, the Bad, and the Transitivity of Better Than *: The Good, the Bad, and the Transitivity of Better Than. *Noûs*, 52(4), 874–899. <https://doi.org/10.1111/nous.12198>

Nebel, J. M. (2019). An Intrapersonal Addition Paradox. *Ethics*, 129(2), 309–343.

<https://doi.org/10.1086/700085>

Nebel, J. M. (2022). Totalism without Repugnance. In J. M. Nebel, *Ethics and Existence* (pp. 200–231). Oxford University Press. <https://doi.org/10.1093/oso/9780192894250.003.0009>

- Nebel, J. M., & Stefánsson, H. O. (2023). Calibration dilemmas in the ethics of distribution. *Economics and Philosophy*, 39(1), 67–98. <https://doi.org/10.1017/S0266267121000298>
- Ng, Y.-K. (1989). What Should We Do About Future Generations?: Impossibility of Parfit's Theory X. *Economics and Philosophy*, 5(2), 235–253. <https://doi.org/10.1017/S0266267100002406>
- Nietzsche, F. W. (2006). *The gay science* (Dover ed). Dover Publications.
- Nover, H., & Hájek, A. (2004). Vexing Expectations. *Mind*, 113(450), 237–249. <https://doi.org/10.1093/mind/113.450.237>
- Nozick, R. (2012). *Anarchy, state, and utopia* (Nachdr.). Blackwell.
- Otsuka, M. (2012). Prioritarianism and the Separateness of Persons. *Utilitas*, 24(3), 365–380. <https://doi.org/10.1017/S0953820812000064>
- Otsuka, M. (2018). How it makes a moral difference that one is worse off than one could have been. *Politics, Philosophy & Economics*, 17(2), 192–215. <https://doi.org/10.1177/1470594X17731394>
- Ottonelli, V. (2010). What Does the Discursive Paradox Really Mean for Democracy? *Political Studies*, 58(4), 666–687. <https://doi.org/10.1111/j.1467-9248.2010.00821.x>
- Owens, D., Horrocks, J., & House, A. (2002). Fatal and non-fatal repetition of self-harm: Systematic review. *British Journal of Psychiatry*, 181(3), 193–199. <https://doi.org/10.1192/bjp.181.3.193>
- Parfit, D. (1982). Future Generations: Further Problems. *Philosophy & Public Affairs*, 11(2), 113–172. JSTOR.
- Parfit, D. (1986). *Reasons and Persons* (1st ed.). Oxford University Press Oxford. <https://doi.org/10.1093/019824908X.001.0001>
- Parfit, D. (1991). *Equality or Priority*. University of Kansas, Department of Philosophy.
- Parfit, D. (2004). Overpopulation and the Quality of Life. In T. Tännsjö & J. Ryberg (Eds.), *The Repugnant Conclusion* (Vol. 15, pp. 7–22). Springer Netherlands. https://doi.org/10.1007/978-1-4020-2473-3_2

- Parfit, D. (2012). Another Defence of the Priority View. *Utilitas*, 24(3), 399–440.
<https://doi.org/10.1017/S095382081200009X>
- Parfit, D. (2016). Can We Avoid the Repugnant Conclusion? *Theoria*, 82(2), 110–127.
<https://doi.org/10.1111/theo.12097>
- Parfit, D. (2017a). Future People, the Non-Identity Problem, and Person-Affecting Principles: Future People, the Non-Identity Problem, and Person-Affecting Principles. *Philosophy & Public Affairs*, 45(2), 118–157. <https://doi.org/10.1111/papa.12088>
- Parfit, D. (2017b). *On what matters*. Oxford university press.
- Pettigrew, R. (2019). *Choosing for changing selves* (First edition). Oxford University Press.
- Pettigrew, R. (forthcoming). Should longtermists recommend hastening extinction rather than delaying it? *The Monist*.
- Pettit, P. (2001). Deliberative Democracy and the Discursive Dilemma. *Philosophical Issues*, 11(1), 268–299. <https://doi.org/10.1111/j.1758-2237.2001.tb00047.x>
- Pivato, M. (2014). Additive representation of separable preferences over infinite products. *Theory and Decision*, 77(1), 31–83. <https://doi.org/10.1007/s11238-013-9391-2>
- Pivato, M. (2018). Rank-Additive Population Axiology. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.3231904>
- Plato. (1975). *Philebus* (J. C. Gosling, Ed.). Clarendon Pr.
- Portmore, D. W. (1999). Does the Total Principle have any Repugnant Implications? *Ratio*, 12(1), 80–98. <https://doi.org/10.1111/1467-9329.00078>
- Pressman, M. (2015). A Defence of Average Utilitarianism. *Utilitas*, 27(4), 389–424.
<https://doi.org/10.1017/S0953820815000072>
- Pummer, T. (2013). Intuitions about large number cases. *Analysis*, 73(1), 37–46.
<https://doi.org/10.1093/analys/ans134>
- Pummer, T. (2017). *Lopsided Lives* (Vol. 1). Oxford University Press.
<https://doi.org/10.1093/oso/9780198808930.003.0014>

- Pummer, T. (2018). Spectrum arguments and hypersensitivity. *Philosophical Studies*, 175(7), 1729–1744. <https://doi.org/10.1007/s11098-017-0932-3>
- Pummer, T. (2022). Sorites on What Matters. In T. Pummer, *Ethics and Existence* (pp. 498–524). Oxford University Press. <https://doi.org/10.1093/oso/9780192894250.003.0019>
- Pummer, T. (2023). *The Rules of Rescue: Cost, Distance, and Effective Altruism*. Oxford University Press.
- Pummer, T. (manuscript). *The Requiring/Justifying Solution to the Non-Identity Problem*.
- Qizilbash, M. (2007). The Mere Addition Paradox, Parity and Vagueness*. *Philosophy and Phenomenological Research*, 75(1), 129–151. <https://doi.org/10.1111/j.1933-1592.2007.00063.x>
- Qizilbash, M. (2018). On Parity And The Intuition Of Neutrality. *Economics and Philosophy*, 34(1), 87–108. <https://doi.org/10.1017/S0266267117000281>
- Rabinowicz, W. (1995). To Have One’s Cake and Eat It, Too: Sequential Choice and Expected-utility Violations: *Journal of Philosophy*, 92(11), 586–620. <https://doi.org/10.2307/2941089>
- Rabinowicz, W. (2000). Money pump with foresight. In M. J. Almeida (Ed.), *Imperceptible harms and benefits*. Kluwer Academic Publishers.
- Rabinowicz, W. (2008). Value Relations. *Theoria*, 74(1), 18–49. <https://doi.org/10.1111/j.1755-2567.2008.00008.x>
- Rabinowicz, W. (2009a). Broome and the Intuition of Neutrality. *Philosophical Issues*, 19, 389–411. JSTOR.
- Rabinowicz, W. (2009b). Incommensurability and Vagueness. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 83, 71–94. JSTOR.
- Rabinowicz, W. (2012). Value Relations Revisited. *Economics and Philosophy*, 28(2), 133–164. <https://doi.org/10.1017/S0266267112000144>
- Rabinowicz, W. (2022a). Can Parfit’s Appeal to Incommensurabilities in Value Block the Continuum Argument for the Repugnant Conclusion? In W. Rabinowicz, *Ethics and*

Existence (pp. 430–460). Oxford University Press.

<https://doi.org/10.1093/oso/9780192894250.003.0016>

Rabinowicz, W. (2022b). Getting Personal: The Intuition of Neutrality Reinterpreted. In G.

Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (pp. 113–141). Oxford University Press.

<https://doi.org/10.1093/oxfordhb/9780190907686.013.17>

Rabinowicz, W. (2022c). Personalized Neutral-Range Utilitarianism with Incommensurable Lives

– What Form Does It Take? And Is It Repugnant? *Filosofiska Notiser*, 9(1), 63–86.

Rachels, S. (1998). Is it Good to Make Happy People? *Bioethics*, 12(2), 93–110.

<https://doi.org/10.1111/1467-8519.00098>

Rachels, S. (2001). A Set of Solutions to Parfit's Problems. *Noûs*, 35(2), 214–238.

<https://doi.org/10.1111/0029-4624.00294>

Rachels, S. (2004). Repugnance Or Intransitivity: A Repugnant But Forced Choice. In T. Tännsjö &

J. Ryberg (Eds.), *The Repugnant Conclusion: Essays on Population Ethics* (pp. 163–186).

Springer Netherlands. https://doi.org/10.1007/978-1-4020-2473-3_9

Rawls, J. (1972). *A theory of justice*. Clarendon Press.

Rawls, J. (2003). *A theory of justice* (Rev. ed., 5.-6. printing). Belknap Press of Harvard Univ.

Press.

Raz, J. (1986). *The morality of freedom*. Clarendon Press ; Oxford University Press.

Rendall, M. (2015). Mere Addition and the Separateness of Persons. *The Journal of Philosophy*,

112(8), 442–455. JSTOR.

Rice, C. M. (2013). Defending the Objective List Theory of Well-Being. *Ratio*, 26(2), 196–211.

<https://doi.org/10.1111/rati.12007>

Riley, J. (1993). On Quantities and Qualities of Pleasure. *Utilitas*, 5(2), 291–300.

<https://doi.org/10.1017/S0953820800005811>

Riley, J. (1999). Is Qualitative Hedonism Incoherent? *Utilitas*, 11(3), 347–358.

<https://doi.org/10.1017/S0953820800002557>

- Riley, J. (2008). Millian Qualitative Superiorities and Utilitarianism, Part I. *Utilitas*, 20(3), 257–278. <https://doi.org/10.1017/S0953820808003142>
- Riley, J. (2009). Millian Qualitative Superiorities and Utilitarianism, Part II. *Utilitas*, 21(2), 127–143. <https://doi.org/10.1017/S0953820809003434>
- Roberts, E., & Wasserman, D. (2017). Dividing and Conquering the Non-Identity Problem. In S. M. Liao & C. O’Neil (Eds.), *Current controversies in bioethics*. Routledge, Taylor & Francis Group.
- Roberts, M. A. (1998). *Child versus childmaker: Future persons and present duties in ethics and the law*. Rowman & Littlefield.
- Roberts, M. A. (2003). Can it Ever Be Better Never to Have Existed At All? Person-Based Consequentialism and a New Repugnant Conclusion. *Journal of Applied Philosophy*, 20(2), 159–185. <https://doi.org/10.1111/1468-5930.00244>
- Roberts, M. A. (2009). The nonidentity problem and the two envelope problem: When is one act better for a person than another? In M. A. Roberts & D. T. Wasserman (Eds.), *Harming future persons: Ethics, genetics and the nonidentity problem* (pp. 201--228). Springer.
- Roberts, M. A., & Wasserman, D. T. (Eds.). (2009). *Harming future persons: Ethics, genetics and the nonidentity problem*. Springer.
- Rosenqvist, S. (2020). *Hedonistic Act Utilitarianism. Action Guidance and Moral Intuitions*. Uppsala University, Department of Philosophy.
- Ross, J. (2015). Rethinking the Person-Affecting Principle. *Journal of Moral Philosophy*, 12(4), 428–461. <https://doi.org/10.1163/17455243-01204004>
- Ross, W. D., & Stratton-Lake, P. (2002). *The right and the good* (New ed.). Clarendon Press.
- Roussos, J. (2021). Normative Formal Epistemology as Modelling. *The British Journal for the Philosophy of Science*, 718493. <https://doi.org/10.1086/718493>
- Roussos, J. (2022). Modelling in Normative Ethics. *Ethical Theory and Moral Practice*, 25(5), 865–

889. <https://doi.org/10.1007/s10677-022-10326-4>
- Ryberg, J. (1996a). Is The Repugnant Conclusion Repugnant? *Philosophical Papers*, 25(3), 161–177. <https://doi.org/10.1080/05568649609506547>
- Ryberg, J. (1996b). Parfit's Repugnant Conclusion. *The Philosophical Quarterly*, 46(183), 202. <https://doi.org/10.2307/2956387>
- Ryberg, J. (2004). The Repugnant Conclusion And Worthwhile Living. In T. Tännsjö & J. Ryberg (Eds.), *The Repugnant Conclusion* (Vol. 15, pp. 239–255). Springer Netherlands. https://doi.org/10.1007/978-1-4020-2473-3_13
- Sartre, J.-P. (1973). *Existentialism and humanism*. Eyre Methuen.
- Savage, L. J. (1972). *The foundations of statistics* (2d rev. ed). Dover Publications.
- Scanlon, T. (2000). What we owe to each other (First Harvard University Press paperback edition). The Belknap Press of Harvard University Press.
- Schoenfield, M. (2014). Decision making in the face of parity. *Philosophical Perspectives*, 28(1), 263–277. <https://doi.org/10.1111/phpe.12044>
- Schopenhauer, A. (2020). *The Cambridge edition of the works of Schopenhauer. [World] Volume 2: The world as will and representation / Arthur Schopenhauer ; translated and edited by Judith Norman, Alistair Welchman, Christopher Janaway ; with an introduction by Christopher Janaway* (J. Norman, A. Welchman, & C. Janaway, Trans.; First paperback edition). Cambridge University Press.
- Schopenhauer, A., Norman, J., Welchman, A., & Janaway, C. (2010). *The world as will and representation*. Cambridge University Press.
- Scovronick, N., Budolfson, M. B., Dennig, F., Fleurbaey, M., Siebert, A., Socolow, R. H., Spears, D., & Wagner, F. (2017). Impact of population growth and population ethics on climate change mitigation policy. *Proceedings of the National Academy of Sciences*, 114(46), 12338–12343. <https://doi.org/10.1073/pnas.1618308114>
- Sebo, J. (2023). The Rebugnant Conclusion: Utilitarianism, Insects, Microbes, and AI Systems.

Ethics, Policy & Environment, 26(2), 249–264.

<https://doi.org/10.1080/21550085.2023.2200724>

Segall, S. (2022). Quarantined Prioritarianism. In S. Segall, *Ethics and Existence* (pp. 552–576).

Oxford University Press. <https://doi.org/10.1093/oso/9780192894250.003.0021>

Seidenfeld, T. (1995). When normal and extensive form decisions differ. In *Studies in Logic and the Foundations of Mathematics* (Vol. 134, pp. 451–463). Elsevier.

[https://doi.org/10.1016/S0049-237X\(06\)80056-X](https://doi.org/10.1016/S0049-237X(06)80056-X)

Shaffer, D. (1996). Psychiatric Diagnosis in Child and Adolescent Suicide. *Archives of General Psychiatry*, 53(4), 339. <https://doi.org/10.1001/archpsyc.1996.01830040075012>

Sider, T. R. (1991). Might theory X be a theory of diminishing marginal value? *Analysis*, 51(4), 265–271. <https://doi.org/10.1093/analys/51.4.265>

Sidgwick, H. (1981). *The methods of ethics*. Hackett Pub. Co.

Skorupski, J. (1999). *Ethical explorations*. Oxford university press.

Smuts, A. (2013). Five Tests for What Makes a Life Worth Living. *The Journal of Value Inquiry*, 47(4), 439–459. <https://doi.org/10.1007/s10790-013-9393-x>

Smart, J. J. C., & Williams, B. (2008). *Utilitarianism for and against* (27. printing). Cambridge Univ. Press.

Snedegar, J. (2017). *Contrastive reasons* (First edition). Oxford University Press.

Snedegar, J. (2021). Reasons, competition, and latitude. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics Volume 16* (1st ed.). Oxford University Press.

<https://doi.org/10.1093/oso/9780192897466.001.0001>

Sophocles. (401 C.E.). *Oedipus at Colonus* (E. Grennan & R. Kitzinger, Trans.; 2005th ed.). Oxford University Press.

Sobel, D. (2002). Varieties of Hedonism. *Journal of Social Philosophy*, 33(2), 240–256.

<https://doi.org/10.1111/0047-2786.00007>

Spears, D., & Budolfson, M. (2019). Why Variable-Population Social Orderings Cannot Escape the

Repugnant Conclusion: Proofs and Implications. *SSRN Electronic Journal*.

<https://doi.org/10.2139/ssrn.3468615>

Spears, D., & Budolfson, M. (2021). Repugnant conclusions. *Social Choice and Welfare*, 57(3), 567–588. <https://doi.org/10.1007/s00355-021-01321-2>

Stefánsson, H. O. (Forthcoming). Longtermism and social risk-taking. In H. Greaves, D. Thorstad, & J. Barret (Eds.), *Essays on Longtermism*. Oxford University Press.

Sumner, L. W. (1996). *Welfare, happiness, and ethics*. Clarendon.

Stroppa, L. (2019). *Choosing who lives our life*. <https://doi.org/10.14275/2465-2334/20199.STR>

Stroppa, L. (2021). Vite Sempre Degne Di Essere Vissute, Vite Mai Degne Di Essere Vissute e Lotterie per la Vita. *Bioetica: Rivista Interdisciplinare*, 3/2021, 429–452.

Stroppa, L. (2023). Plant Ethics and Climate Change. In G. Pellegrino & M. Di Paola (Eds.), *Handbook of the Philosophy of Climate Change* (pp. 899–917). Springer International Publishing. https://doi.org/10.1007/978-3-031-07002-0_73

Suominen, K., Isometsä, E., Suokas, J., Haukka, J., Achte, K., & Lönnqvist, J. (2004). Completed Suicide After a Suicide Attempt: A 37-Year Follow-Up Study. *American Journal of Psychiatry*, 161(3), 562–563. <https://doi.org/10.1176/appi.ajp.161.3.562>

Tännsjö, T. (1998). *Hedonistic utilitarianism*. Edinburgh University Press.

Tännsjö, T. (2002). Why We Ought to Accept the Repugnant Conclusion. *Utilitas*, 14(3), 339–359. <https://doi.org/10.1017/S0953820800003642>

Tannsjö, T. (2002). Why We Ought to Accept the Repugnant Conclusion. *Utilitas*, 14(3), 339–359. <https://doi.org/10.1017/S0953820800003642>

Tännsjö, T. (2020). Why Derek Parfit had reasons to accept the Repugnant Conclusion. *Utilitas*, 32(4), 387–397. <https://doi.org/10.1017/S0953820820000102>

Temkin, L. S. (1987). Intransitivity and the Mere Addition Paradox. *Philosophy & Public Affairs*, 16(2), 138–187. JSTOR.

Temkin, L. S. (1996). A Continuum Argument for Intransitivity. *Philosophy & Public Affairs*,

25(3), 175–210. JSTOR.

Temkin, L. S. (2012). *Rethinking the good: Moral ideals and the nature of practical reasoning*. Oxford university press.

Temkin, L. S. (2015). Rethinking Rethinking the Good. *Journal of Moral Philosophy*, 12(4), 479–538. <https://doi.org/10.1163/17455243-01204006>

Thomas, T. (2018). Some Possibilities in Population Axiology. *Mind*, 127(507), 807–832. <https://doi.org/10.1093/mind/fzx047>

Thomas, T. (2022a). Are Spectrum Arguments Defused by Vagueness? *Australasian Journal of Philosophy*, 100(4), 743–757. <https://doi.org/10.1080/00048402.2021.1920622>

Thomas, T. (2022b). Separability and Population Ethics. In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (pp. 270–295). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190907686.013.21>

Thomas, T. (2023). The asymmetry, uncertainty, and the long term. *Philosophy and Phenomenological Research*, 107(2), 470–500. <https://doi.org/10.1111/phpr.12927>

Thomson, J. J. (1997). The Right and the Good. *The Journal of Philosophy*, 94(6), 273. <https://doi.org/10.2307/2564542>

Thornley, E. (2021). The impossibility of a satisfactory population prospect axiology (independently of Finite Fine-Grainedness). *Philosophical Studies*, 178(11), 3671–3695. <https://doi.org/10.1007/s11098-021-01621-4>

Thornley, E. (2022). Critical Levels, Critical Ranges, and Imprecise Exchange Rates in Population Axiology. *Journal of Ethics and Social Philosophy*, 22(3). <https://doi.org/10.26556/jesp.v22i3.1593>

Thornley, E. (Unpublished). *A Non-Identity Dilemma for Person Affecting Theories*.

Tiberius, V. (2013). Beyond the experience machine: How to build a theory of well-being. In M. Haug (Ed.), *Philosophical Methodology: The Armchair or the Laboratory?* (0 ed., pp. 408–425). Routledge. <https://doi.org/10.4324/9780203798997-32>

- Timmer, D. (2022). Justice, Thresholds, and the Three Claims of Sufficiencyarianism*. *Journal of Political Philosophy*, 30(3), 298–323. <https://doi.org/10.1111/jopp.12258>
- Tucker, C. (2023). Parity, moral options, and the weights of reasons. *Noûs*, 57(2), 454–480. <https://doi.org/10.1111/nous.12410>
- Vallentyne, P. (1991). The Problem of Unauthorized Welfare. *Noûs*, 25(3), 295. <https://doi.org/10.2307/2215505>
- Vallentyne, P. (2009). Broome on Moral Goodness and Population Ethics. *Philosophy and Phenomenological Research*, 78(3), 739–746. <https://doi.org/10.1111/j.1933-1592.2009.00261.x>
- Van Der Lugt, M. (2021). *Dark Matters: Pessimism and the Problem of Suffering*. Princeton University Press. <https://doi.org/10.1515/9780691226156>
- Velleman, J. D. (1991). Well-being and time. *Pacific Philosophical Quarterly*, 72(1), 48–77. <https://doi.org/10.1111/j.1468-0114.1991.tb00410.x>
- Venkatesh, N. (2020). Repugnance and Perfection. *Philosophy & Public Affairs*, 48(3), 262–284. <https://doi.org/10.1111/papa.12165>
- Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior* (60. anniversary ed., 4. printing, and 1. paperback printing). Princeton University Press.
- Voorhoeve, A. (2008). 'Heuristics and biases in a purported counter-example to the acyclicity of "better than"'. *politics, philosophy & economics*, 7(3), 285-299.
- Voorhoeve, A., & Fleurbaey, M. (2016). Priority or Equality for Possible People? *Ethics*, 126(4), 929–954. <https://doi.org/10.1086/686000>
- Whewell, W. (1971). *Mathematical exposition of some doctrines of political economy*. A. M. Kelley.
- Wilkinson, H. (2022). In Defense of Fanaticism. *Ethics*, 132(2), 445–477. <https://doi.org/10.1086/716869>
- Williams, B. (1995). *Making Sense of Humanity: And Other Philosophical Papers 1982–1993* (1st

- ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511621246>
- Williams, B. (1981). *Moral Luck: Philosophical Papers 1973–1980* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139165860>
- Williamson, P. (2021). A New Argument Against Critical-Level Utilitarianism. *Utilitas*, 33(4), 399–416. <https://doi.org/10.1017/S0953820821000133>
- Winsor, J. (2017). *Christopher Columbus and How He Received and Imparted the Spirit of Discovery* (fifth edition). HANSEBOOKS.
- Wolf, S. (1997). Happiness and Meaning: Two Aspects of the Good Life. *Social Philosophy and Policy*, 14(1), 207–225. <https://doi.org/10.1017/S0265052500001734>
- Wolf, S. (2010). *Meaning in Life and Why It Matters*. Princeton University Press. <https://doi.org/10.1515/9781400834594>
- Wolf, S. (2016). Meaningfulness: A Third Dimension of the Good Life. *Foundations of Science*, 21(2), 253–269. <https://doi.org/10.1007/s10699-014-9384-9>
- Woods, D. (2014). *Schopenhauer's pessimism*. University of Southampton.
- Woodward, J. (1986). The Non-Identity Problem. *Ethics*, 96(4), 804–831. JSTOR.
- Zuber, S., Venkatesh, N., Tännsjö, T., Tarsney, C., Stefánsson, H. O., Steele, K., Spears, D., Sebo, J., Pivato, M., Ord, T., Ng, Y.-K., Masny, M., MacAskill, W., Lawson, N., Kuruc, K., Hutchinson, M., Gustafsson, J. E., Greaves, H., Forsberg, L., ... Asheim, G. B. (2021). What Should We Agree on about the Repugnant Conclusion? *Utilitas*, 33(4), 379–383. <https://doi.org/10.1017/S095382082100011X>