ESSAYS IN THE ETHICS OF DISTRIBUTION

KORBINIAN RÜGER

Balliol College, University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2020

Abstract

This thesis is about the allocation of benefits under scarcity constraints. In five standalone papers, it provides answers to some aspects of the following central question of distributive ethics: Out of a set of distributions of well-being over a group of individuals, which distribution should we choose?

The first of the five papers draws a connection between the moral notions of interpersonal aggregation and equality. It argues that aggregation sceptics should be egalitarians. The second paper puts forward a novel model of "limited aggregation" that is not subject to the same objections that have recently been raised against other such models. The third paper criticizes one of the most influential views of cashing out non-aggregative ethics under conditions of risk, "ex ante contractualism." The fourth paper defends an ex post approach as an alternative to ex ante contractualism. The fifth and final paper tries to incorporate some of our common-sense moral intuitions into a coherent theory of variable population ethics. It does so by combining reasons stemming from an impersonal axiology with reasons stemming from the complaints people can raise against our choices.

This thesis consists of 58,915 words.

Acknowledgments

I owe thanks to many people. First of all, I would like to thank my supervisors. Hilary Greaves provided many penetrating, often challenging, but always constructive comments on almost all aspects of this thesis. The same goes for Ralf Bader, who saved me from many errors. They both made this thesis much better.

Johann Frick, Jeff McMahan, Andreas Mogensen, and Tom Sinclair served as supervisors, examiners or advisors during some of my time on the DPhil. Alex Voorhoeve supervised my MSc thesis at the LSE in which some of the work in this thesis originates. I am very lucky to have worked with all of them.

Many more people helped me with some aspects of this thesis. They include Jessica Fischer, Joe Horton, Kacper Kowalczyk, Benjamin Lange, Michal Masny, Johanna Privitera, Theron Pummer, Bastian Steuwer, Patrick Tomlin, and audiences at the University of Oxford, UCL, Princeton University, the University of Ghent, LMU Munich, Cardiff University, the University of St. Andrews, the University of Salzburg, and Karsruhe Institute of Technology. I thank all of them.

For generous financial support, I want to express my sincere gratitude to Stiftung der Deutschen Wirtschaft, the German-American Fulbright Commission, Balliol College, the Philosophy Faculty in the University of Oxford, the Thomas und Ulla Kolbeck Stiftung, and the Society for Applied Philosophy.

The by far greatest debt is owed to my parents, Barbara Berendt-Rüger and Matthias Rüger. Throughout my life they have unconditionally supported me and my projects in every way imaginable. My sister, Alexandra Rüger, has played a similar role. Everything was built on that foundation. For this, I am forever grateful.

Finally, I thank Isabella Wolf for her patience, love, and support, especially during the final stretches of writing this thesis.

Danke.

For my parents

Contents

Introduction

Out of a set of distributions of well-being over a group of individuals, which distribution should we choose? This DPhil thesis consists of five standalone papers exploring the prospects of a broadly non-consequentialist approach to answering this central question of distributive ethics. I believe that this question is closely linked to the evaluative task of determining the best outcome from a set of outcomes, i.e. to finding the correct axiology. It is often the case, that once we have identified the best outcome, it follows that we ought to choose that outcome.

However, I don't believe that this is always the case. For example, if realizing the best outcome would require my violating someone's right or if doing so would be grossly unfair, it may be the case that I should not, all things considered, realize that outcome. This is why I believe that the evaluative task of determining the best outcome and the normative question of which outcome we ought to choose can come apart and why my thesis is an exercise in non-consequentialist ethics. I draw together different strands of moral thought and put them to work in three central areas of contemporary moral philosophy: interpersonal aggregation, moral decision making under risk, and variable population ethics.

Although the individual papers can be read on their own, there are common threads connecting them. In this introduction, I first provide a brief overview of the papers and, second, point out the connections between them. I will also mention a few of the many research questions this thesis leaves unaddressed, but that I hope to get to in future work.

Synopsis

The first paper, "Aggregation and Equality," examines the connections between two hitherto largely unrelated notions in moral philosophy, first, the interpersonal aggregation of harms and benefits and, second, the value of equality in well-being. In this paper, I present two arguments for the claim that moral theory should be fully aggregative. I show that these arguments can most plausibly be rejected on egalitarian grounds. This means that sceptics regarding interpersonal aggregation should be egalitarians.

In the second paper, "Aggregation with Constraints," I move to the question of what a plausible account denying full aggregation could look like. Since the rejection of all forms of interpersonal aggregation strikes most people as implausible, some critics of fully aggregative views have proposed theories of *Limited Aggregation*. These occupy the middle ground between fully aggregative and non-aggregative views. Recently, Limited Aggregation has been criticized for having counterintuitive implications that seem even worse than the counterintuitive implications of fully aggregative and non-aggregative views it tried to escape. I here propose a new view of Limited Aggregation that does better than existing accounts in this regard. It is more modest than existing accounts of Limited Aggregation, but it retains the view's core idea. This, I claim, is the thought that sometimes very strong individual claims stand in the way of realizing the best outcome.

With the third paper, "On Ex Ante Contractualism," I move to discussing moral decision making under uncertainty. T.M. Scanlon's contractualism is one of the most influential non-aggregative moral theories of recent years. Johann Frick's *ex ante* contractualism is the best worked out approach to date to extend Scanlonian contractualism to conditions of risk. Ex ante contractualism holds that in situations involving risk, we ought to act in accordance with principles that license the action that satisfies the strongest individual claim, where those claims are a function of the expected value that a given policy gives each person ex ante. I here criticize ex ante contractualism. I argue that adopting it would have far reaching implications that contractualists, or other sceptics concerning interpersonal aggregation, would find very hard to accept.

The fourth paper, "Minimizing the Strongest Complaint Under Risk," adds to this critique and defends an ex post approach as an alternative to ex ante contractualism and to hybrid accounts that hold that we should take an ex ante approach to one kind of case and an ex post approach to another kind of case. According to the ex post approach I favour, it is permissible to choose a prospect P over another prospect Q if and only if the expected strongest complaint against Q is at least as strong as the expected strongest complaint against P. I argue that such an ex post approach is the most attractive way to extend contractualism and other non-aggregative theories to conditions of risk.

With the fifth and final paper, "Impersonal Value, Personal Complaints," I leave the terrain of *fixed* population ethics and move to *variable* population ethics. In this paper I defend a broadly pluralist approach to population ethics that takes into account reasons stemming from an impersonal axiology as well as reasons stemming from the complaints that individual people affected by our choices could put forward against these choices. The account rejects the so-called "Neutrality Intuition," according to which bringing lives worth living into existence, is in itself "morally neutral." However, it salvages a "Weak Asymmetry," according to which we have stronger reasons to prevent miserable lives from coming into existence than we have to bring good lives into existence. Nonetheless, it holds that we *do* have reasons and sometimes an obligation to bring into existence such good lives.

Let me now move on to describe a picture of ethics that underlies much of my thinking and therefore the papers in this thesis.

Common Themes

In the end, what we want to know from normative ethics is what we ought morally to do. We want to know, given a choice between a number of possible acts, which act we ought to choose. That is, we want to know the correct *deontology*. Deontology is concerned with classifying alternatives we can choose among as obligatory, permissible, or prohibited. Given any remotely plausible moral theory, our axiology delivers part of our deontology. If we know that one act α will result in state of affairs A that is strictly better than the state of affairs B an alternative act β results in, then this gives us a prima facie reason in favour of performing α rather than β. And, if there are no other reasons at play, this gives us an obligation to perform α rather than β.[1] Further, if consequentialism is correct, then, roughly, all moral reasons are axiological. However, if consequentialism is not correct, and we should not assume that it is, then there are also non-axiological moral reasons. These could, for example, stem from individual rights that people have, from considerations of justice, or from other specific complaints that people can raise against our choices.

I should note that my use of the term "deontology" here may strike some readers as non-standard. Unlike other contemporary philosophers, I do not take "deontology" to be a distinct moral theory or a set of moral theories ("rule-based ethics"), but rather a necessary part of any moral theory. For example, I do not think that utilitarianism and deontology are fundamentally opposed to each other. Rather, utilitarianism is a specific variant of deontology in that there is only one rule: do whatever maximizes the good!

In my view then, the term "deontology" is best reserved for that part of any moral theory that deals with "the right," i.e. what we ought to do or have reasons to do, and can thus be contrasted with axiology, which is the part of a theory that deals with "the good." In part then, this thesis is about the ways in which morality can depart from promoting the good.

---

[1] This follows from a suitable bridge principle, like the following, which I suggest in chapter 5 of this thesis.

> *From Reason to Obligation.* If you have a reason to choose X over Y, and no reason to choose Y over X, then you have an obligation to choose X over Y.

What I have said so far, suggests a picture of morality that I will not argue for, but that some might object to. This is the idea that, first, we can do work in axiology without worrying about the correct overall moral theory and, second, we can think of the ways in which we can depart from axiological thinking without worrying about what the correct axiology actually looks like.

Such a way of thinking about ethics may strike some as misguided who think that the good can be analysed in terms of the right or, in my terminology, axiology can in some sense be reduced to deontology.[2] On such accounts, the "division of labour" between axiology and broader moral theory that I suggested, may be hard to maintain. This is because, on some such accounts, the good itself does not provide normative reasons. Rather, whatever it is that makes an object or a state of affairs good or valuable, is what provides reasons to take a certain attitude towards this object. These reasons in turn are what makes the object valuable to begin with. In a slogan, the buck stops with reasons and not with value.

I am sceptical about this approach to morality. I *do* think that value in itself provides reasons. However, even if this wasn't the case, I am unsure whether this would count against the picture of morality that I sketched above. Even if axiology could in some sense be reduced to deontology, it still makes sense to talk about axiology without worrying about the underlying moral theory. By analogy, if we think that, for example, biology can be reduced to physics, this does not

[2] Such a view of morality is most often associated with "fitting-attitudes" or "buck-passing" accounts of value. See e.g. A. C. Ewing, *The Definition of Good* (London: Routledge and Kegan Paul, 1947). chapter 2, and T.M. Scanlon, *What We Owe to Each Other* (Cambridge, Mass. ; London: Belknap Press of Harvard University Press, 1998). chapter 1.

mean that we must think that we cannot meaningfully talk about biology without worrying about the correct underlying physical theory.[3] Still, the approach to moral theory that underlies the papers in this thesis assumes that we have reasons to promote the good *because* it is good. For example, it assumes that we have reasons to promote well-being simply because well-being is good.

Does this mean that I have stacked the deck in favour of consequentialism? I do not think so. Admittedly, some may object that the kind of non-consequentialism I examine in this thesis is only a "half-hearted" form of non-consequentialism, a kind of "consequentialism plus constraints," or maybe "consequentialism plus constraints and prerogatives."[4] I think I can live with that description. As I said above, I do happen to think that we have reason to promote or even "maximize" the good, a supposition that, of course, has a consequentialist flavour to it. Nonetheless, I think there are other moral reasons, that are independent from this supposition and which put limits on the extent to which we ought to, or are even permitted, to maximize the good. Whether, for the purposes of this thesis, this makes me a "half-hearted non-consequentialist" or, maybe, a "half-hearted consequentialist," or both, I am unsure. In any case, I am fine with either label. It is this picture of the structure of morality that underlies my arguments in each of the five papers.

---

[3] I thank Andreas Mogensen for discussion of this point.

[4] I take the term "consequentialism plus constraints" from Theron Pummer, "Whether and Where to Give," *Philosophy & Public Affairs* 44, no. 1 (December 2016): 77–95. p. 79. For a criticism of this kind of "half-hearted consequentialism," see Thomas Sinclair, "Are We Conditionally Obligated to Be Effective Altruists?," *Philosophy & Public Affairs* 46, no. 1 (January 2018): 36–59.

Avenues for Future Research

Unavoidably, there are many questions my thesis leaves unanswered. Let me briefly describe two of them that I hope to examine in the future.

First, I want to look into the precise relation between axiological and non-axiological moral reasons. For example, in chapter 2 of this thesis, I put forward the idea that if one person, or a group of people, has a much stronger claim to our aid than any person whose interests conflict with hers, we ought to satisfy this claim. If, however, it is not the case that one person has much more at stake than any other person, we ought to do what realizes the best outcome. The best outcome, in turn, is determined by the most plausible axiology, which might well be fully aggregative.

This suggests one way in which non-axiological reasons could limit the extent to which it is required or even permissible to promote the good. Traditionally, there are broadly two ways to limit the extent to which axiological reasons determine our overall moral theory: side constraints and personal prerogatives. For example, many people think that we have a general duty not to lie. This is a side constraint. However, virtually nobody thinks that this duty holds if, say, the continued existence of the universe depended on violating it. In this case, the axiological stakes would simply be too high. This suggests that side-constraints should not be understood as being absolute.

Now, it seems that especially in the context of population ethics, we often encounter cases where the axiological stakes are very high. This is especially the case when thinking about the long-term future. In such cases then, it seems as if non-axiological moral reasons have little or no role to play anymore if they are not absolute. I raise this worry in chapter 5 of this thesis and hope to come back to it in the future.

The second and related topic I could not do justice in this thesis, but would like to get to in future research, are the practical implications of population ethics. What exactly are the obligations that we have towards people that may live in the future? For example, are we obligated to have children if we knew that they would have lives worth living and all other things are equal? In the last chapter of this thesis, I give an argument for the affirmative answer to this question. However, I remain unsure whether it is correct. Are we obligated to make future people's lives go as well as possible? If so, are these duties of goodness or duties of justice?

These questions are still very much up for grabs. Almost all choices we make today have downstream effects on the identity, the number, and the well-being of future people. Often, these effects are not properly taken into account when decisions are made. There is, of course, a rapidly growing literature in population ethics that studies the moral ramifications of these effects. However, much of population ethics is exclusively concerned with evaluating different possible futures along a single dimension, overall value or goodness.

John Broome, for example, argues that, given the "non-identity problem," to a large extent we cannot have duties of justice towards future people.[5] I am unsure about this and in future research, I want to investigate this claim. It concerns what I call *population justice*. Population justice deals with the specific moral duties that we,

[5] See John Broome, "The Public and Private Morality of Climate Change," *The Tanner Lectures on Human Values*, 2012.

individuals and groups of individuals, owe to future people. Given that the actions we could take to fulfil our duties towards future people, often themselves have effects on the identity of these people (i.e. the non-identity problem), the ties that usually bind together duty holders and those to whom duties are owed, are severed. I think we might be able to replace these ties through a new account of what our duties of justice actually consist in. According to this account, much of moral philosophy would be mistaken in assuming that duties of justice are owed to specific individuals. Rather, the account holds that these duties are owed to *locations*, which are occupied by specific individuals. This setup could open up new ways to think about population justice.

1 | Aggregation and Equality

In ethics, interpersonal aggregation is the combining of claims or the well-being of individual people in order to decide what we morally ought to do. Under utilitarianism, for example, we always ought to do what will lead to the greatest sum of well-being across all individuals. There is an influential strand in contemporary moral philosophy that objects to the aggregative element in utilitarianism and other theories. This kind of scepticism about interpersonal aggregation is a staple of contemporary non-consequentialist moral thought.[1] Take, for example, the following case:

> *Death vs. Headaches:* We can either save Ann from a terminal illness or prevent some number *n* of different people from suffering a mild headache.

The claim that there is a number *n* for which we ought to prevent the headaches rather than save Ann is denied by many people. Rather, they take it for granted that no matter how large *n* is, we ought to save Ann. We must not let one person die just so that we can prevent a number of people from suffering a mild headache, even if this number

[1] This is not to say that *any* non-aggregative or only partially aggregative theory must be non-consequentialist. See e.g. Larry S. Temkin, *Rethinking the Good : Moral Ideals and the Nature of Practical Reasoning* (New York: Oxford University Press, 2015). especially chapters 2 and 3.

is very large. The contrary, however, is implied by the following principle which, as we will see, is hard to reject.

> *Full Aggregation*. For any harms X and Y, where Y is worse than X, there is some *n*, such that we ought to save *n* people from X rather than one different person from Y.

Many sceptics of Full Aggregation are also on record as egalitarians. They are committed to the claim that equality in people's well-being levels is valuable in itself or that we have reasons to choose more equal rather than less equal outcomes. However, usually, issues regarding skepticism about interpersonal aggregation are treated separately from equality. I here argue that there are good reasons to treat them together. I argue that Full Aggregation might most plausibly be rejected on egalitarian grounds. This is because the most powerful positive arguments *for* Full Aggregation rely on premises that conflict with egalitarianism. If one is committed to egalitarianism, one should reject these premises. And since, as I will argue, these are the weakest premises of the arguments, Full Aggregation can most plausibly be resisted by taking on egalitarian commitments.

I further suggest that the proper way to understand these egalitarian commitments is on non-axiological rather than axiological grounds. These are commitments rooted in considerations of fairness. One upshot of this is that we can reject Full Aggregation on deontic grounds but concede that our axiology could be fully aggregative. This would, for example, imply that we sometimes ought to choose suboptimal outcomes if choosing the best outcome would be unfair to some people.

I will proceed as follows. In the first two sections I introduce two arguments, which I take to be representatives of the strongest positive arguments for Full Aggregation. The first is based on Harsanyi's aggregation theorem implying utilitarianism, the second is based on an argument by Alastair Norcross. I show that both of these arguments each rely on premises that could be rejected on egalitarian grounds. In the third section I assess how this egalitarianism could be interpreted.

But first, some preliminaries. Out of a set of distributions of well-being over a fixed group of individuals, which distribution should we choose? This is the guiding question I am interested in for the purposes of this paper. Since all the outcomes I consider, are distributions of well-being, I will use the terms "outcome" and "distribution" interchangeably. Further, I will assume that other than their respective well-being levels the people in all the outcomes do not differ in any morally relevant way. The guiding question is closely linked to the evaluative task of ordering outcomes in terms of better and worse all things considered, i.e. finding the correct axiology. However, though there might be a tight connection, the evaluative question of determining the best outcome and the normative question of which outcome we ought to choose, are not the same. This is important because the arguments I consider in the following have axiological counterparts based on axiological principles. I shall not, however, directly be concerned with these axiological principles.

Further note that it could be argued that the strongest arguments *for* Full Aggregation are actually negative arguments *against* non-aggregative or only partially aggregative theories. These are arguments for the claim that all these theories are implausible so that any plausible theory *must* be fully aggregative.[2] While such negative arguments have merit, they shall not concern me here.

Let me now move on to the first of the two arguments for Full Aggregation, based on John Harsanyi's aggregation theorem.

I. The Argument from Harsanyi's Aggregation Theorem

With Harsanyi's theorem it can be proven that if individual preferences and "social preferences" obey the axioms of expected utility theory and, in addition, social preferences are impartial and satisfy what is known as the Ex Ante Pareto Principle, then our overall social preferences can be represented by the utilitarian value function, which takes the following form and is fully aggregative.[3]

$$U(X) = u_1(X) + u_2(X) + \cdots + u_n(X)$$

Here, X is the outcome in question, $U(X)$ is the overall utility of that outcome and $u_i(X)$ is the personal utility of individual i in that outcome. We can immediately see that this theory satisfies Full Aggregation. For example, u_1 could be arbitrarily small, representing the fact that i_1 dies in outcome X. As long as n is large enough, even the slightest benefit to all these people will outweigh i_1's death.

In line with common practice in economics, Harsanyi frames his argument in terms of individual and social *preferences*. John Broome

---

2 For such an account see Joe Horton, "Always Aggregate," *Philosophy & Public Affairs* 46, no. 2 (March 2018): 160–74. I deal with Horton's argument in chapter 2 of this thesis.

3 See John C. Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63, no. 4 (1955): 309–21.

has argued, convincingly in my view, that we can and should recast the argument in terms of personal and general (i.e. overall) *value*.[4] Though I take my lead from Broome, I will yet again employ the argument in a different, more directly normative, way. I will recast the argument and its premises in terms of what we ought to choose, i.e. in terms of *rational* choiceworthiness in the individual case and *moral* choiceworthiness in the social case. Employed in this way, Harsanyi's theorem is a valid argument for Full Aggregation as stated above. So, anyone wanting to resist Full Aggregation needs to reject one or more of Harsanyi's assumptions recast in this way.

His first assumption is that individual rational choice obeys the axioms of expected utility theory.[5] As a normative theory for individuals, expected utility theory is very attractive. Of course, it is likely to be false as an empirical theory about how individuals actually act, but that is not our topic. So, I shall not question this first assumption here. I will further assume throughout that ideally rational preferences provide an adequate measure of a person's well-being.

The second assumption is much more controversial. This is the assumption that our social preferences should likewise obey expected utility theory. By "social preferences" Harsanyi means the preferences a social planner has regarding possible choices between different distributions of well-being. This implies that if our "social preferences" are given by our moral theory, as I assume they should be, then social

---

[4] See John Broome, "General and Personal Good," in *The Oxford Handbook of Value Theory*, ed. Iwao Hirose and Jonas Olson (Oxford University Press, 2015), 249–66.
[5] There are, of course, different versions of expected utility theory. However, Harsanyi's theorem works with many of them. Harsanyi himself relied on Marshak's theory. See Jacob Marschak, "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica* 18, no. 2 (1950): 11–141.

preferences are whatever our moral theory tells us about these kinds of choices. Given that the scope of this paper does not go beyond these kinds of choices, we can equate "social preference" with "moral theory" for our purposes.[6]

I will not go over all the standard axioms of expected utility theory, but here focus on

> *The Sure Thing Principle*. If two prospects P and Q yield the same outcome in some state of nature, then our choice between P and Q depends only on the outcomes in other states of nature.[7]

Let me explain this principle with a simple example.

	S1 (p=0.5)		***S2 (p=0.5)***	
	i	***j***	***i***	***j***
A	1	0	1	0
B	1	0	0	1

Diamond's Choice

Here we are choosing between two prospects, A and B. There are two equiprobable states of nature, S1 and S2, and two individuals, *i* and *j*. A and B lead to the same distribution in S1. So, according to the Sure-Thing Principle our preference between A and B should depend only on what happens in S2. Assuming that we should be indifferent

[6] In other words, all the choices considered here are choices between different ways of conferring benefits on a group of individuals. There are no murders, maimings, or other atrocities involved that might be ruled out on independent grounds.

[7] This principle is sometimes also called "Strong Independence" or "Strong Separability". See e.g. chapter 5 of John Broome, *Weighing Goods* (Cambridge, Mass.: Basil Blackwell, 1991).

between the two possible distributions in S2,[8] the Sure Thing Principle implies that we should be indifferent between A and B.

The principle seems intuitively attractive. If S1 obtains, we would be indifferent given that both outcomes are the same, and if S2 obtains we would also be indifferent, given that both outcomes are symmetric and we do not favour either of *i* and *j*. It, therefore, seems that we should also be indifferent between prospects A and B.

Nonetheless, some will be inclined to say that we should not be indifferent, but that we ought to choose B and thus reject the Sure Thing Principle. As Peter Diamond writes, ”[B] seems strictly preferable to me, since it gives [*j*] a fair shake while [A] does not.”[9] In other words, equality in expected well-being matters, according to Diamond. This is the first point at which we can see that an appeal to equality could serve as a roadblock on the route to Full Aggregation. If the Sure Thing Principle is unacceptable because it violates equality, then Harsanyi’s aggregation theorem does not establish Full Aggregation.[10]

I now move on to the final assumption of Harsanyi’s. This is the following principle.

---

[8] This assumption is backed by a weak *Anonymity* condition.

> *Weak Anonymity*. If two outcomes A and B differ only in that the identities of two people have been permuted, we should be indifferent between A and B.

[9] Peter A. Diamond, “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment,” *Journal of Political Economy* 75, no. 5 (1967): 765–66.

[10] There are possibilities to establish very similar conclusions to Harsanyi’s that do not rely on the Sure-Thing principle. See e.g. Marc Fleurbaey, “Two Variants of Harsanyi’s Aggregation Theorem,” *Economics Letters* 105, no. 3 (December 2009): 300–302. These, however, rely on dominance principles that likewise seem to violate the sort of ex ante equality that Diamond had in mind.

> *Strong Ex Ante Pareto*. If prospect P has higher expected value than prospect Q for some *i* and Q does not have higher expected value than P for any *i*, then we ought to choose P over Q.

Strong Ex Ante Pareto is intuitively very attractive. If expected utility theory is the correct theory of individual rationality under conditions of uncertainty, as I have granted above, then denying Ex Ante Pareto would sometimes require us to act against the best interest of everyone involved. If they are rational by the lights of the theory, they will maximize their expected well-being. So, if they could choose for themselves, they would adhere to Ex Ante Pareto and they would want us to adhere to it when choosing on their behalf. This means that rejecting Ex Ante Pareto requires us to override the consent (explicit or implicit) of the very people on whose behalf we are choosing. This is a large bullet to bite. Nonetheless, Ex Ante Pareto can be questioned and has been questioned, on the basis of equality. Take, for example, the following sort of case.[11]

	S1 (p=0.5)		***S2 (p=0.5)***	
	i	***j***	***i***	***j***
A	2+ε	0	0	2+ε
B	1	1	1	1

Myerson's Choice

For a sufficiently small ε, many people will here say that we ought to choose B. This could be, for example, because they say that the extra overall value in A is not worth giving up on the equality in B. By

[11] To my knowledge, Roger Myerson was the first to introduce such a case. See Roger B. Myerson, "Utilitarianism, Egalitarianism, and the Timing Effect in Social Choice Problems," *Econometrica* 49, no. 4 (July 1981): 883. See also Broome, *Weighing Goods*. p. 185.

choosing B, we can make sure that the resulting distribution will be perfectly equal. A on the other hand, guarantees that our distribution will be unequal. Still, according to Ex Ante Pareto we ought to choose A. It gives both, *i* and *j*, higher expected utility than B.

Here then, we have another clash between Harsanyi's theorem and possible considerations stemming from worries about the importance of equality. These considerations, however, are markedly different from those that led Diamond and others to doubt the Sure-Thing Principle. Those were worries about equality in *expected* well-being. Here we are concerned with worries about equality in *final* well-being. If we care about this kind of *outcome-* or *ex post* equality, we should reject Ex Ante Pareto. If we care about *ex ante* equality, we should reject the Sure-Thing Principle.[12] Either way of insisting on the importance of equality will likely lead one to reject the argument for Full Aggregation based on Harsanyi's theorem.

I will explore the question of which the more plausible route is in section three. For now, let us move on to the second argument for Full Aggregation.

[12] Of course, we could also care about both kinds of equality and reject both. However, since I find these principles attractive, I think we should try to violate few rather than more of them.

II. The Argument from Individual Risk Taking

In presenting this second argument for Full Aggregation, I take my lead from Alastair Norcross.[13] Norcross's argument is specifically geared at denying the following claim, which Norcross calls

> *Worse*. Other things being equal, it is worse that one person die a premature death than that *any number* of people suffer moderate headaches for twenty-four hours.

This, of course, is just the evaluative analogue of denying Full Aggregation. Norcross's goal is to provide an argument against Worse. For the most part, Norcross limits his discussion to saving people from deaths and headaches, but there is no principled reason not to generalize his argument to capture all harms and benefits, which is what I will do. Also, as before, I will be concerned with normative analogues of the axiological claims made by Norcross.

The argument's starting point is the supposition that the number of people affected by our choices matter. For example, when we have to choose between saving one person from death and saving five different people from death, we ought to save the five just because they are greater in number or, as a general principle,

---

[13] See Alastair Norcross, "Great Harms from Small Benefits Grow: How Death Can Be Outweighed by Headaches," *Analysis* 58, no. 2 (1998): 152–158. Thomas Dougherty in Thomas Dougherty, "Aggregation, Benificence and Chance," *Journal of Ethics & Social Philosophy. Vol. 7, No. 2*, 2013. presents essentially the same argument. He claims, however, that his is different since "Norcross's argument presupposes a consequentialist approach to the ethics of risk and aggregation. Norcross's premises and conclusion are claims about which outcomes are better or worse than other outcomes […]" (p. 4). While this latter claim is true, it is straightforward to employ normative analogues of Norcross's evaluative premises, which is the strategy I will employ.

> *Numbers Matter*. For every harm X, there is some *n*, such that we ought to save *n* people from X rather than one different person from X.

Note that unlike the other principles considered so far, Numbers Matter, as the name suggests, itself introduces a very limited dose of interpersonal aggregation. I take it however, that it is a requirement every plausible moral theory needs to fulfil. Nonetheless, the literature on how to ground this claim is enormous and, of course, there are some dissenting voices.[14] I will leave this debate to the side and hope that Numbers Matter is plausible enough to be believed without much argument. I will just note that as stated, this is a very modest claim since, in principle, *n* could be quite large. I myself doubt that it has to be. In fact, I am very confident that *n* is two. But we do not need that much stronger claim.

The second premise we need is that we sometimes ought to save people from a risk of harm rather than a different person, or a smaller group of people, from suffering that harm with certainty. This is straightforward and intuitively obvious in cases where we can, for example, either save one person from death or ten different people from a very high chance of death, say 99.9%.[15] It is very likely that by saving the ten from the risk, and thereby letting the one die, we prevent ten deaths rather than one.[16] And since, as I assumed above, Numbers Matter, this is what we ought to do. Things are intuitively less clear

---

[14] See e.g. John M. Taurek, "Should the Numbers Count?," *Philosophy & Public Affairs* 6, no. 4 (1977): 293–316. and G. E. M. Anscombe, "Who Is Wronged? Philippa Foot on Double Effect: One Point," in *Human Life, Action and Ethics: Essays* (Exeter, UK: Imprint Academic, 2005), 249–51. Chapter 19.
[15] Assume these individual probabilities are independent from each other.
[16] This chance is greater than 99%.

when the risks are smaller. Suppose we can either save one from certain death or ten different people from a 10% risk of death. Here it is no longer clear and a lot of factors, like our attitude to risk, are involved. But it seems clear that as the number of people we can save from the risk grows, at some point, we ought to help them and let the one die. Suppose for example there are one billion people facing the 10% risk. Then it is clear that we ought to help them, even if that means that we let the one die. This suggests the following general principle.

> *Many Person Risk*. For every harm X and every probability p, there is some n, such that we ought to save n people from suffering X with probability p, rather than one different person from X.

Some will be tempted to deny this principle even if they accept Numbers Matter. One reason, again, could stem from equality. If, as in the context of the Sure Thing Principle, one cares about equalizing expected utility, there will be cases where one might think that doing so requires rejecting Many Person Risk. Take a case where we can either save one from death or any number of people from a tiny risk of death. Then, no matter how small that risk is, Many Person Risk says that there is a number of people such that we ought to lower their risk to 0, rather than save the one. On a plausible measure of equality, however, we would make the expected utility distribution more equal by saving the one, that is lowering her risk from 100% to 0, rather than lowering everyone else's risk to 0.

Johann Frick, for example, argues that in many such situations as the above, we should satisfy the strongest *ex ante claim*, where this claim is a function of the expected utility of the person who has the largest stake in our situation. However, even Frick recognizes that if n gets very large, we can no longer plausibly adhere to his choice rule.[17] We should not save one person if that comes at the cost of millions of expected deaths. Again, this weaker claim is all we need.

The next step in the argument is to accept expected utility theory for individual choices, specifically when choosing on someone else's behalf. I have already claimed in the context of discussing Harsanyi's theorem, that I will not doubt the appropriateness of choosing according to expected utility theory in individual choice contexts. Still, since the role this claim plays in the argument of this section is different, it is worth spending some time on it.

> *Individual Rationality.* If prospect P has higher expected value than prospect Q for individual i, then we ought to choose P over Q on i's behalf.

This just means that when choosing on someone else's behalf, we ought to choose in the same way this person would choose for herself if perfectly rational according to expected utility theory.[18] One important part of expected utility theory is a Continuity axiom. Applied to our context it is as follows.

---

[17] Johann Frick, "Contractualism and Social Risk," *Philosophy & Public Affairs* 43, no. 3 (June 2015): 175–223.

[18] I am here bracketing thorny issues relating to personal autonomy and consent. However, I am attracted to the idea that in situations in which we lack information about the person's actual preferences, i.e. about how they themselves would choose in a certain situation, we should choose for them as if they were perfectly rational according to expected utility theory, e.g. that they are risk neutral with respect to well-being. For a defense of this claim see e.g. Michael Otsuka, "Prioritarianism and the Measure of Utility," *Journal of Political Philosophy* 23, no. 1 (March 2015): 1–22.

> *Individual Continuity*. For any individual *i* and outcomes X, Y and Z, where X is better for *i* than Y, which is better for *i* than Z, there is a (non-trivial) probability p, such that we ought to choose prospect pX (1-p)Z over Y when choosing on *i*'s behalf.

The intuitive idea behind this principle is very simple: Often it is worth taking risks of harm in order to get even small benefits. In our day to day lives, we do this all the time. Whenever we get in a car, take out our bike, or cross the street, we accept a non-negligible risk of injury or even death in order to incur benefits that, for the most part, are quite small. Individual Continuity says that we can apply this sort of thinking to any case exhibiting this structure and choose accordingly on someone else's behalf. For example, it is plausible that we ought to alleviate someone's headache by administering a pain killer that carries a very slight chance of deadly side effects if this chance is small enough. We often would not judge it irrational to take such a painkiller in order to alleviate one's headache and we do not judge it immoral when someone like a doctor chooses to administer it to a patient or a parent to their child. The next principle we need just extends this thinking to a group of people.

> *Weak Ex Ante Pareto.* If prospect P has higher expected value than prospect Q for all *i*, then we ought to choose P over Q.

This, of course, is just a weaker and therefore even more plausible version of Strong Ex Ante Pareto employed in Harsanyi's argument above. In effect, this just says that if we can choose according to expected utility theory on one person's behalf, then we can do so on behalf of a group of people each facing the same situation as the individual in the one-person case. As already seen in the preceding section, this principle is of great intuitive appeal. But like Harsanyi's stronger Ex Ante Principle, this principle can likewise be questioned on egalitarian grounds. Recall the above choice in Table 2. Here it is again.

	S1 (p=0.5)		***S2 (p=0.5)***	
	i	***j***	***i***	***j***
A	2+ε	0	0	2+ε
B	1	1	1	1

Myerson's Choice

By Weak Ex Ante Pareto, we ought to choose A, but for a small enough ε many people will think that we ought to choose B on egalitarian grounds. How plausible this is, will be the topic of the next section.

But first, we now have all principles in place to complete the second argument for Full Aggregation. For example, let p be 1/1,000,000 and let n be one billion. Then, Many Person Risk says that we ought to save one billion people from a 1/1,000,000 risk of death, rather than one from certain death. This seems very reasonable. By doing so, we expect to save a thousand lives. Of course, it is very unlikely that we will save exactly a thousand lives. But the chance that we save more than ten lives, for example, is approaching 100%. That seems all we need to know. This judgement is backed by Numbers Matter.

Now, let prospect P be death with probability p and life without headache with probability 1-p. Let "prospect" Q be life with headache for certain. Assuming that the expected value of P exceeds the expected value of Q for all i, Weak Ex Ante Pareto implies that we ought to choose P in a case where n people face that choice. That is, we ought to save, for example, one billion from a headache rather than one billion from a 1/1,000,000 risk of death. If the "ought to _ rather than _" relation is transitive in this case, as it seems to be, then it follows that we ought to save one billion from a headache rather than one from death. This generalizes to Full Aggregation.

A few words about transitivity. I write that the "ought to _ rather than _" relation seems to be transitive *in this case*, because it is not clear whether this relation is transitive in all cases. I *am* assuming the "better than" relation to be transitive in all cases.[19] But betterness is not our topic here. There clearly seem to be cases in which "ought to _ rather than _" is not transitive. For example, suppose you have promised your friend to pick him up from the airport. You can now either choose to pick him up or stay at home and watch TV. Suppose, however, there is a third option: outside your house is a burning car with one person inside. So, you could instead of going to the airport or watching TV also risk your life and save the person. Doing so is supererogatory. When we break down this situation into three binary choices, you may permissibly choose to watch TV (and probably calling 911) over risking your life to save the person, which you may permissibly choose over picking up your friend. It is not the case, however, that you can permissibly choose to watch TV over picking up your friend.[20] This is a violation of transitivity.

However, it is certainly not the case that transitivity *never* holds for the "ought to _ rather than _" relation. Claiming this would be plainly

---

[19] Pace e.g. Larry S. Temkin, "Intransitivity and the Mere Addition Paradox," *Philosophy & Public Affairs*, 1987, 138–187.
[20] For other such examples, see e.g. Pummer, "Whether and Where to Give." p. 83 and Shelly Kagan, *The Limits of Morality* (Oxford University Press, 1989). p. 16.

absurd. It holds in most cases. We can say that this relation is *presumptively transitive* and anyone offering us a case in which they claim it is not transitive, has to provide us with good reasons why this is so in this specific case. In my case above, these reasons are that in the choice between watching TV and rescuing a person, you have an agent-centred prerogative to prioritize your well-being over that of a stranger and you have an obligation to keep your promise to your friend unless something of greater importance releases you from this obligation. Of course, one could doubt these reasons for violating transitivity in this case, but they are not obviously implausible. Regarding the role transitivity plays in the above argument for Full Aggregation, I cannot think of any such reasons. Unless someone shows us otherwise, we can safely assume transitivity to hold in this case.

With this route closed, to resist this argument for Full Aggregation, we need to either reject Many Person Risk or Weak Ex Ante Pareto. In the next section I will examine which of these options is more attractive. I will argue that the most promising route for sceptics of Full Aggregation is to reject Weak Ex Ante Pareto.

III. Which Equality?

In the preceding two sections, we have seen that both arguments for Full Aggregation considered here, can be questioned on egalitarian grounds. The argument from Harsanyi's aggregation theorem relies on the Sure Thing Principle and on Strong Ex Ante Pareto. The argument from individual risk taking relies on Weak Ex Ante Pareto and on what I called Many Person Risk. If we are concerned about equalizing ex ante chances of receiving a benefit, we should reject the Sure Thing Principle and plausibly also Many Person Risk. If we are concerned about equalizing actual benefits, we should reject the Ex Ante Pareto principles. So, either way of interpreting egalitarianism in the face of risk, makes it possible for egalitarians to reject two strong arguments for Full Aggregation. This is the central observation of this essay.

In the following, I will sketch a proposal regarding the kind of egalitarianism that I think aggregation sceptics should adopt. By sketching this proposal, I can only scratch the surface of egalitarian thinking. There are many important questions that I need to leave aside. Let me, however, briefly mention one debate which has dominated much of the discourse in egalitarian thinking in recent years, the debate between prioritarianism and "true" egalitarianism.[21]

Roughly, prioritarians care about equality only insofar as they think that, morally speaking, benefits to a person matter more the worse off that person is in absolute terms. Egalitarians, on the other hand, think that equality in well-being matters in itself. Though they agree in many cases, they are fundamentally different theories. Egalitarianism is a *relational* view, prioritarianism is not.[22]

In the following, I ignore prioritarianism. There are two reasons for this. First, I do not find prioritarianism particularly compelling.[23]

---

[21] For a good overview of this recent debate, see Michael Otsuka and Alex Voorhoeve, "Equality Versus Priority," in *The Oxford Handbook of Distributive Justice*, ed. Serena Olsaretti, vol. 1 (Oxford University Press, 2018).
[22] See e.g. John Broome, "Equality versus Priority: A Useful Distinction," *Economics and Philosophy* 31, no. 02 (July 2015): 219–28.
[23] For objections against prioritarianism that I find convincing, see e.g. Michael Otsuka and Alex Voorhoeve, "Why It Matters That Some Are Worse off than Others: An Argument against the Priority View," *Philosophy & Public Affairs* 37, no. 2 (2009): 171–199. and Toby Ord, "A New Counterexample to

Second, and more importantly for the purposes of this essay, it is not clear how adopting prioritarianism can help with rejecting Full Aggregation. Under the most common interpretation, prioritarianism is itself a fully aggregative view. It applies a strictly concave and strictly increasing transformation function to individual well-being levels in order to arrive at the priority weighted moral value of these well-being levels and then sums these values across all individuals.[24] Of course, there are also fully aggregative interpretations of egalitarian views, but it is much less obvious that these interpretations of egalitarianism are the most plausible ones.

Both egalitarianism and prioritarianism are in a position to reject some of the principles that serve as premises of the two above arguments for Full Aggregation. However, egalitarianism has a clear rationale for doing so, whereas prioritarianism violates these principles "by accident." Take the case of Ex Ante Pareto. As we have seen, it is clear why egalitarians who care about equalizing final well-being would want to reject Ex Ante Pareto in some cases. Prioritarians, however, also need to reject Ex Ante Pareto in cases in which it is not at all clear why one would want to do so. For example, they need to reject it even in single-person cases and in some cases where there is no inequality

---

Prioritarianism," *Utilitas* 27, no. 03 (September 2015): 298–302. See also Roger Crisp, "Equality, Priority, and Compassion," *Ethics* 113, no. 4 (2003): 745–763. pp. pp. 750-755. Note that Crisp argues that we should account for our distributive intuitions neither in an egalitarian nor in a prioritarian, but in a *sufficientarian* way. According to this approach, what we should care about is that all have *enough*, i.e. are *sufficiently well-off.* For a prominent statement of this view see Harry Frankfurt, "Equality as a Moral Ideal," *Ethics* 98, no. 1 (1987): 21–43. For a reply, see Paula Casal, "Why Sufficiency Is Not Enough," *Ethics* 117, no. 2 (2007): 296–326. This debate is one of the many issues I need to leave aside.

[24] See e.g. Hilary Greaves, "Antiprioritarianism," *Utilitas* 27, no. 01 (March 2015): 1–42. p. 8.

in final well-being.[25] Egalitarianism on the other hand, does not have these odd implications and has a clear rationale for rejecting some of the above principles, like Ex Ante Pareto.

However, there is still the question about which kind of egalitarianism to adopt. In particular, I am interested in two choice points. First, there is a choice between a concern for equality in *expected* well-being and a concern for equality in *final* well-being. Second, there is a choice between what I call *axiological* egalitarianism and *non-axiological* egalitarianism. In the following, I argue that egalitarians should care about equality of well-being rather than expected well-being and that this concern for equality should be non-axiological rather than axiological.

This is the kind of egalitarianism I favour. However, I want to stress that nothing of what I say in the following will serve as a knock-down argument against one or the other kind of egalitarianism. For example, one could also defend a version of egalitarianism that combines a concern for equality in expected well-being with a concern for equality in final well-being.[26] This would likewise get one out of the two above arguments for Full Aggregation by denying *both* Ex Ante Pareto and the Sure Thing Principle. However, since I find these principles attractive, I think we should look for ways to violate fewer rather than more of them.

---

[25] See e.g. Greaves. p. 37.

[26] See e.g. Alex Voorhoeve and Marc Fleurbaey, "Egalitarianism and the Separateness of Persons," *Utilitas* 24, no. 03 (September 2012): 381–98.

Expected or Final Well-Being?

The first question I want to raise is why we should even think that a concern for equality requires equalizing *expected* well-being. The answer seems obvious: because we care about equality of well-being! But, as John Broome writes, "[a]n expected utility is not at all the same sort of thing as a utility, so what applies to one does not necessarily apply to the other."[27] For example, one might argue that expected well-being is void of intrinsic value in the same way that you cannot eat the expectation of a sandwich. The sandwich is good for you if you eat it and maybe the expectation of it is good for you because of the benefits of anticipation. But this, of course, does not mean that there is anything *intrinsically* valuable about expected well-being. Moreover, if equality in expected well-being *were* good for people, that should already be reflected in the well-being levels in a choice like Diamond's.

Rather, I suggest that expected well-being is only valuable in the sense that maximizing expected well-being will on average maximize well-being. This is why it is plausible that for each individual, maximizing expected well-being is the rational course of action. But it is by no means obvious that expected well-being is the sort of thing to which moral notions like justice or fairness normally apply, that expected well-being is something that we should care about distributing fairly.

There is, however, one way in which one might think that we should care about equalizing expected well-being or benefits. This is in cases in which we are to distribute an *indivisible* good and different people

---

[27] John Broome, "Uncertainty and Fairness," *The Economic Journal* 94, no. 375 (September 1984): 626. See also David Wasserman, "Let Them Eat Chances: Probability and Distributive Justice," *Economics and Philosophy* 12, no. 01 (1996): 29–49.

have equal claims to receive that good. In these cases, it seems that we cannot distribute the good fairly.[28] What we can do, however, is to use randomization and equalize chances of receiving that good. Some take this to be a sort of *equal surrogate satisfaction of claims* in cases in which we cannot satisfy actual claims equally.[29] As John Broome writes, "[e]qual chances provide a surrogate equality in satisfaction."[30]

Suppose we have to decide whether to save Ann or Bob from death. We cannot save both. Further suppose that Ann and Bob are equally non-responsible for their plight, equally deserving and so on. To many, it seems natural in such a case that we ought to, for example, flip a fair coin to decide whom to save. This and similar ideas have also found their way into public policy. For example, some indivisible goods, like exemption from military conscription or access to important medical resources, like access to dialysis machines, have been assigned by lot.[31] Many people seem to think that this is the fairest way to go about making such decisions. But why? Kornhauser and Sager, for example, write the following:

> Suppose we have a single good which is indivisible and two claimants of the good who each enjoy equal and robust moral entitlements to it. An equiprobable lottery has the

[28] Of course, we could refuse to distribute or destroy the good in question. That would be, in some sense, perfectly fair. However, I assume that in most cases when the stakes are high enough, this gain in fairness would be outweighed by other considerations.

[29] See e.g. John Broome, "Fairness," in *Proceedings of the Aristotelian Society*, vol. 91, 1990, 87–101., John Broome, "Kamm on Fairness," Philosophy and Phenomenological Research 58, no. 4 (1998): 955–61. and Lewis. A. Kornhauser and Lawrence G. Sager, "Just Lotteries," Social Science Information 27, no. 4 (1988): 483–516.

[30] Broome, "Kamm on Fairness." p. 956.

[31] For these and other examples, see e.g. John Elster, *Solomonic Judgments: Studies in the Limitations of Rationality* (Cambridge University Press, 1989). pp. 45-102.

> unique capacity to divide the good probabilistically, making possible the allocation of a [50%] chance of receiving the good to each candidate.[32]

However, as Tim Henning writes, "[o]n the face of it, this is a puzzling idea. How can an indivisible good be divisible after all, and what is it to divide a good probabilistically?"[33] For example, in the above case, it is not as if we save the loser of our coin flip to *some degree*. If a good is indivisible, then it is impossible to divide it. That is the whole point. Either Ann dies and Bob lives, or Bob lives and Ann dies. Whatever we do, we will give rise to a manifestly unfair situation. The reasoning embodied in the above quote seems to suggest that this were not the case or, at least, that by flipping a coin we could somehow remedy the tragedy of this manifestly unfair situation. We can't.

To be sure, I am not saying that the use of a fair lottery in these cases is never appropriate. I just want to cast doubt on the idea that the reasons why a lottery might be appropriate in these cases, have, at their core, to do with considerations of equality or fairness. For example, we might want to employ a lottery to prevent bias or even perceived bias. Or, we might want to use a lottery, because it makes things easier, psychologically speaking, for the decision maker. These are all perfectly good reasons, but they are not ultimately grounded in equality. In any case, such considerations are assumed away in a case like Diamond's Choice.

Moreover, if the thought is that equality in expected utility can mitigate necessary inequality in utility, one might think that there is

---

[32] Kornhauser and Sager, "Just Lotteries." p. 491.
[33] Tim Henning, "From Choice to Chance? Saving People, Fairness, and Lotteries," *Philosophical Review* 124, no. 2 (April 2015): 169–206. p. 171. In the following, I am indebted to Henning's discussion of these issues.

value to equalizing chances whenever we can. So, whenever we arrive at an unequal outcome after equalizing chances, we might want to equalize again and again, by adding one coin flip after another. That seems absurd. To take this point even further, we might think that if equalizing (epistemic) chances makes things fairer, we might have a reason to deprive ourselves of information we could easily obtain or already have. That seems absurd. Rather, I submit, what egalitarians should care about in the face of risk is *expected equality in well-being*, not *equality in expected well-being*, given that what they ultimately care about is *equality in well-being*. This implies that egalitarians should not reject the Sure Thing Principle, but Ex Ante Pareto.

However, there is still the question *why* we should reject Ex Ante Pareto. Should we do so because equality contributes to the value of an outcome, or because equality is important in some other, non-axiological way? I next turn to this question.

Equality as Fairness

To the extent that we would care about equality in well-being, *why* is it that we should care about that equality? There are broadly two lines that we can take in answering this question. The first is that we should care about equality because it by itself makes an outcome better. I shall refer to this option as *axiological* egalitarianism. The second is that we should care about equality even though it by itself does not make an outcome better, but for some other reason. Unsurprisingly, I shall refer to this as *non-axiological* egalitarianism.[34] This second approach holds

[34] Parfit draws a similar distinction between teleological or "telic egalitarianism" and deontological or "deontic egalitarianism." See Derek

that we have reasons to care about outcome equality that do not stem from the goodness of outcomes themselves. One such account holds that we should care about equality if and when it is a sign of fairness, a notion I already alluded to above. For example, as Larry Temkin writes:

> If I give one piece of candy to Andrea, and two to Rebecca, Andrea will immediately assert 'unfair!' This natural reaction suggests an intimate connection between equality and fairness. Arguably, concern about equality is that portion of our concern about comparative fairness that focuses on how people fare relative to others. Specifically, concern about equality reflects the view that inequality is bad when, and because, it is unfair, where the unfairness consists in one person being worse off than another no more deserving.[35]

I agree with Temkin that we should care for equality out of a concern about (relative) fairness.[36] However, I disagree with Temkin when it comes to how we should interpret the unfairness generated by inequality in well-being. Temkin interprets it axiologically. I think that this is a mistake. It is very controversial that inequality on its own can make an outcome worse, even if we think that equality matters in

---

Parfit, "Equality and Priority," *Ratio (New Series)* X, no. 3 (1997): 202–21. Although the telic/deontic distinction is well known, I shall use the terms axiological and non-axiological to refer to this distinction. This is because I find Parfit's terminology misleading, because axiological egalitarianism, of course, has deontic significance through whatever function or theory connects the axiological to the deontic.

[35] Larry Temkin, "Equality as Comparative Fairness," *Journal of Applied Philosophy* 34, no. 1 (February 2017): 43–60. p. 45.

[36] Of course, much remains to be said for and against this claim. For example, we might still doubt whether fairness understood in Temkin's way is, at its core, morally significant. Roger Crisp, for example, suggests a debunking explanation that ultimately grounds fairness in envy. Crisp writes that "generalized envy may have become, through a process of cultural evolution, the principle that it is bad if, through no fault of his or her own, one individual does worse than another." Crisp, "Equality, Priority, and Compassion." p. 749.

principle. Inequality might make an outcome worse through harming those who are suffering from the inequality and are thus experiencing unfairness. That is, equality might be axiologically significant, but not above and beyond its contributions to how well the lives of specific people are going.[37]

Rather, I think we should understand fairness as grounded in the individual claims that people have on us to make them as well off as we can. These individual claims have to be balanced against each other. This way of caring about equality is deeply individualistic and rooted in the idea that what we should do when choosing between outcomes is to take note of each person's perspective and compare these perspectives one by one. This echoes Thomas Nagel's account of equality, who writes that "the general form of moral reasoning is to put yourself into other people's shoes." [38] For Nagel, this requires that to be right, any distribution must in principle be acceptable from each person's standpoint. This amounts to a criterion of universal acceptability or unanimity. From this it follows that, prima facie, any good should be distributed equally among the members of a group.[39] Only this distribution will be acceptable for each person. However,

> [w]here there is conflict of interests, no result can be acceptable to everyone. But it is possible to assess each result from each point of view to try to find the one that is least unacceptable to the person to whom it is most unacceptable. This means that any other alternative will be more unacceptable to someone than this alternative is to

[37] This is similar to John Broome's view about the matter. See Chapter 9 of Broome, *Weighing Goods*.
[38] Thomas Nagel, "Equality," in *Mortal Questions* (Cambridge: Cambridge University Press, 1979), 106–28. p. 126.
[39] Note, however, that some interpret Nagel in a *prioritarian* rather than an egalitarian way. See e.g. Crisp, "Equality, Priority, and Compassion." p. 752. and Parfit, "Equality and Priority."

> anyone. The preferred alternative is in that sense the least unacceptable, considered from each person's point of view separately.[40]

So, when claims compete, they ought to be satisfied in a way that echoes the prima facie commitment to equality. According to this approach, outcome inequality matters if and when we could have made the relatively worse-off better off. So, the genesis of an outcome matters. We cannot just look at the outcome and judge whether the inequality in it is objectionable or not. In addition to its intuitive plausibility, I take this non-axiological approach to be more attractive, because, as I shall briefly argue, unlike axiological egalitarianism it is not (obviously) vulnerable to two well-known objections to egalitarianism, the *Levelling Down Objection* and the *Scope Objection*.[41]

The Levelling Down Objection goes as follows. If inequality makes an outcome worse in itself, as axiological egalitarianism has it, removing that inequality by making the better off worse off must make things better in some respect. Since such "levelling down" does not make things better at all, axiological egalitarianism is false.

I myself am not convinced by this objection. If inequality *does* actually make an outcome worse, then, of course, removing that inequality would make the outcome better.[42] It is by no means obvious that a view that runs into the levelling down objection must be false. However,

---

[40] Nagel, "Equality." p. 123.

[41] For both objections see Parfit, "Equality and Priority." For the Scope Objection also see Parfit's objection against average utilitarianism based on the "irrelevance of Egyptology" in Derek Parfit, *Reasons and Persons* (Oxford University Press, 1986): 420. For the Levelling Down objections, see also Dennis McKerlie, "Egalitarianism," *Dialogue* 23 (1984): 223–37.

[42] See e.g. Larry S. Temkin, "Equality, Priority or What?," *Economics and Philosophy* 19, no. 1 (April 2003): 61–87. pp. 67-68.

many people take this objection to be decisive and look for other ways to account for our egalitarian intuitions.[43]

Since non-axiological egalitarians do not believe that equality by itself makes an outcome better, they are not committed to the claim that removing inequality through levelling down makes an outcome better and are thus not vulnerable to (this reading of) the Levelling Down Objection. The above sketched rationale, that considerations of equality are rooted in the fact that individuals have claims on us to make them as well off as possible, is immune to the Levelling Down objection. After all, people have claims that we make them better off, not that we make others worse off. So, bringing down the best-off people will not make things fairer under this interpretation. And since on this account, all that is valuable about equality is ultimately grounded in fairness, this sort of equality is nothing that should move us.

The second objection, the Scope Objection, in turn, goes as follows. When we engage in evaluating different outcomes, we are in effect comparing different possible worlds. To assess the overall value of these possible worlds, we need to take into account the well-being of people everywhere and at all times. And if equality between people adds to the goodness of an outcome and inequality subtracts from the goodness of an outcome, then we need to take account of the equality between people everywhere and at all times. This would, for example, imply that when deciding whether to have a child, one factor we might need to consider is how well off our child would be compared to people that lived thousands of years ago. If our child would be

[43] See e.g. Parfit, "Equality and Priority."

considerably better off than these people, then that would be one reason not to have the child. Since this is absurd, axiological egalitarianism must be false.

Again, on a plausible construal of non-axiological egalitarianism, this objection does not apply. If, as sketched above, equality is important if and when it is fair, and fairness is rooted in the claims that people have on us to make them as well off as possible, we need not take into account the well-being of people that are long dead, since dead people do not have these sorts of claims on us. To be sure, future people might. But taking into account the well-being of future people when deciding whether to have a child is far less absurd than taking into account the well-being of past people, since future people might very well be affected by our choices. Thus construed, non-axiological egalitarianism is not vulnerable to the Scope Objection.

This concludes my brief discussion of different kinds of egalitarianism. I have argued that to resist Full Aggregation we can take one egalitarian commitments. Furthermore, this sort of egalitarianism should be understood as caring about equality in *final* rather than *expected* well-being and should be construed non-axiologically.

Let me briefly note, however, that even if one cares about equality in *expected* well-being, this too is best understood non-axiologically. For if Diamond is correct and we should equalize ex ante chances, it is clear that that has nothing to do with the claim that equal outcomes are better than unequal outcomes. Rather, it is a claim about procedural fairness. Diamond does not dispute the claim that whatever we do, the outcome we will bring about is equally as good as any other we could have brought about. However, the way we arrive at the outcome matters too, according to Diamond. Equalizing chances is fairer. So, this way of denying Harsanyi's argument relies on the notion that equality and fairness matter, but not on the notion that equality matters for our axiology. Either way, we are better off understanding egalitarianism non-axiologically.

Conclusion

Of course, the nature of the value equality is an important and very complicated issue that I have only been able to scratch the surface of. But what I have shown in this essay is that aggregation sceptics can plausibly take on egalitarian commitments to resist Full Aggregation. There is, however, the further question what a plausible theory eschewing Full Aggregation could look like. I turn to this question in the next chapter.

2 | Aggregation with Constraints[*]

Suppose that in

> *Death vs. Headaches:* We can either save Ann from a terminal illness or prevent some number of different people from suffering a mild headache.

Many people think that there is no number of people suffering from a mild headache for which we ought to help them rather than save Ann from death. Next, suppose that in

> *Death vs. Paraplegias:* We can either save Ann from a terminal illness or prevent some number of different people from suffering permanent paraplegia.

Many of the same people think that whether we ought to save Ann from death depends on the number of people we could save from paraplegia instead. These two judgements seem to be in tension. In *Death vs. Paraplegias* these people allow that a larger number of lesser harms (the paraplegias) together outweigh a smaller number of greater harms (the one death) in importance.[1] In *Death vs. Headaches,* these same people deny that preventing the headaches can ever be more

[*] A slightly different version of this chapter is forthcoming under the same title in *Utilitas*.

[1] If you think that paraplegia is worse than death, then simply swap one for the other.

important than saving Ann from death, no matter how many people would actually suffer the headache. According to these two judgements, interpersonal aggregation of harms determines what we ought to do in *Death vs. Paraplegias* but does not determine what we ought to do in *Death vs. Headaches*. Views exhibiting this pattern of judgements are committed to *Limited Aggregation.* Limited Aggregation occupies the middle ground between *Full Aggregation* and *Non-Aggregation.* [2]

> *Full Aggregation*. For any harms X and Y, where Y is worse than X, there is some *n*, such that we ought to save *n* people from X rather than one different person from Y.

> *Non-Aggregation*. For any harms X and Y, where Y is worse than X, there is no *n*, such that we ought to save *n* people from X rather than one different person from Y.

One fully aggregative view is utilitarianism. According to this view, we ought to do whatever leads to the best outcome, where the best outcome is the one maximizing overall well-being. The overall well-being of an outcome, in turn, is simply an additive, unweighted function of all the individual well-being levels in that outcome. Strictly

[2] For a recent systematic account of Limited Aggregation that sparked a lively debate see Alex Voorhoeve, “How Should We Aggregate Competing Claims?,” *Ethics* 125, no. 1 (October 2014): 64–87. For accounts in support of the Limited Aggregation intuitions preceding Voorhoeve’s see e.g. Dale Dorsey, “Headaches, Lives and Value,” *Utilitas* 21, no. 01 (March 2009): 36–58. or F. M. Kamm, “Nonconsequentialism,” in *The Blackwell Guide to Ethical Theory*, ed. Hugh LaFollette and Ingmar Persson, 2nd ed. (Blackwell Publishing, 2013). For rare examples of strictly non-aggregative views, see e.g. Anscombe, “Who Is Wronged? Philippa Foot on Double Effect: One Point.” and Taurek, “Should the Numbers Count?” For fully aggregative views see e.g. Alastair Norcross, “Comparing Harms: Headaches and Human Lives,” *Philosophy & Public Affairs* 26, no. 2 (1997): 135–67. and Derek Parfit, “Justifiability to Each Person,” *Ratio (New Series)* XVI, no. 4 (2003): 368–90.

aggregative views have no problem explaining the common intuition in *Death vs. Paraplegias*. For some number, the benefits to the people we can save from paraplegia, if we do save them, will together outweigh the benefit we can give to Ann. These views, however, cannot explain the common intuition in *Death vs. Headaches*.[3]

Non-aggregative views have the opposite problem. They cannot account for the judgement that it is required, or even permissible, to save some number of people from paraplegia, rather than Ann from death. One family of non-aggregative views holds that individuals we could benefit, have "claims" to these benefits or "complaints" against not getting a benefit that do not lend themselves to the kind of combining together as required by aggregative views. Rather, we ought to respect the "separateness of persons" and, after conducting a series of pairwise comparisons, do whatever satisfies the strongest individual claim, or equivalently, minimizes the strongest individual complaint.[4] On these views, we therefore ought to save Ann from death rather than any number of people from paraplegia. This judgement is at least as implausible as the strictly aggregative judgement that we ought to spare a large number of people from a headache rather than save Ann from death.

The fact that Limited Aggregation makes neither of these intuitively implausible judgements gives us some reason to believe that such a

[3] This is not to say that there *could not* exist any aggregative view that could resist letting Ann die in Death vs. Headaches. We could, for example have a view with a fully aggregative, but lexicographic axiology, according to which well-being consists of, say, two elements, where no amount of one can ever outweigh any amount of the other. Such views shall not concern me here, however.

[4] See e.g. Nagel, "Equality." and Scanlon, *What We Owe to Each Other*. pp. 189-248.

view is true. It is my aim in this essay to spell out what I take to be the most plausible version of Limited Aggregation. I am not offering a full justification of the view. Rather, I want to explore the most plausible way to hold on to the two intuitions in *Death vs. Headaches* and *Death vs. Paraplegias*, if we so want to. As I will argue, we then should adopt the view that if one person has much more at stake than any person whose interests conflict with hers, we ought to do whatever serves the interests of that person. Otherwise, we ought to do what realizes the best outcome, as determined by the most plausible axiology which might well be fully aggregative. This is a much more modest proposal than existing accounts of Limited Aggregation. But as we will see, it retains its core commitments without suffering from the same defects as existing accounts.

I will proceed in five steps. Building on a proposal by Alex Voorhoeve, in the first section, I will introduce Limited Aggregation in more detail. In the second section, I deal with an objection to Limited Aggregation that has recently been brought forward by Patrick Tomlin. This objection holds that the view faces insurmountable problems in cases where individuals are added to the groups of people we can help. In response to this objection, in the third section, I introduce my own approach to Limited Aggregation, which I call *Aggregation with Constraints*, and argue that it does not suffer from the same problems. In the fourth section, I deal with a recent objection to Limited Aggregation by Joe Horton. As I will uncover, Horton's objection is an instance of the more general, but hitherto unstated, worry that Limited Aggregation is implausible because it violates a certain *Separability* condition. According to this condition, roughly, our choice between distributions should not depend on elements that are the same in these distributions. I will argue that this should not surprise us, as views of Limited Aggregation are by definition relational. Whether a given claim can determine what we ought to do, depends on how strong this claim is compared to other claims under consideration. A violation of Separability is thus a *central feature of*, rather than an *objection to* the view. In the fifth and final section, I briefly deal with the objection that my Aggregation with Constraints fails in cases where we have more than two options available.

I. Limited Aggregation

The intuitive idea behind Limited Aggregation is that interpersonal aggregation of harms and benefits should sometimes make a difference to what we ought to do and sometimes not. In *Death vs. Headaches* it should not make a difference, in *Death vs. Paraplegias* it should. Recent accounts of Limited Aggregation can be traced back to comments T.M. Scanlon made in *What we Owe to Each Other.* While justifying the "individualist restriction" of his Contractualism, which, in principle, forbids any form of interpersonal aggregation, he offers the following qualification:

> If one harm, though not as serious as another, is nonetheless serious enough to be morally 'relevant' to it, then it is appropriate, in deciding whether to prevent more serious harms at the cost of not being able to prevent a greater number of less serious ones, to take into account the number of harms involved on each side. But if one harm is not only less serious than, *but not even relevant to*, some greater one, then we do not need to take the number of people who would suffer these two harms into account

> in deciding which to prevent, but should always prevent the more serious harm.[5]

Derek Parfit suggested a view that according to him captured the essence of Scanlon's remarks. He calls it the

> *Close Enough View*. When [and only when] burdens to different people are close enough in size, one greater burden could be morally outweighed by a sufficient number of lesser burdens.[6]

This view allows us to capture the aforementioned pair of intuitions about Deaths vs. Headaches and Death vs. Paraplegias. It was recently spelled out in a systematic way by Alex Voorhoeve.[7] On his approach, which he calls

> *Aggregate Relevant Claims* (*ARC*).
>
> 1. Each individual whose well-being is at stake has a claim on you to be helped. […]
>
> 2. Individuals' claims compete just in case they cannot be jointly satisfied.
>
> 3. An individual's claim is stronger the more her well-being would be increased by being aided. […]
>
> 4. A claim is relevant if and only if it is sufficiently strong relative to the strongest competing claim.
>
> 5. You should choose an alternative that satisfies the greatest sum of strength-weighted, relevant claims.[8]

---

[5] My emphasis. It is in fact very hard to allow for even Limited Aggregation within Scanlon's Contractualism. (see e.g. Michael Otsuka, "Saving Lives, Moral Theory, and the Claims of Individuals," *Philosophy & Public Affairs* 34, no. 2 (2006): 109–135.). This, however, shall not be my topic here.

[6] Parfit, "Justifiability to Each Person.": 278.

[7] See Voorhoeve, "How Should We Aggregate Competing Claims?"

[8] Voorhoeve. p. 66. I have amended the view to reflect a simplifying assumption that the strength of a person's claim only turns on the size of a given benefit and not the absolute well-being level of a person.

Voorhoeve explains that, in a pairwise comparison between two claims, the weaker claim is relevant to the stronger claim if and only if the person holding that weaker claim would not be required to give it up if she herself were the one to make the choice. The approach thus takes into account the subjective viewpoint of individuals from which it seems permissible to favour oneself to a certain degree. For example, it seems permissible for me to save myself from paraplegia rather than a stranger from death, but impermissible to spare myself a minor headache at the cost of the death of a stranger.

In *Death vs. Headaches,* ARC says that we ought to save Ann irrespective of the number of people we could spare the headache because the individual claims to be spared a headache are not close enough to Ann's claim to be saved from death. We, therefore, ought to satisfy the single strongest individual claim. In *Death vs. Paraplegias* on the other hand, the individual claims to be saved from paraplegia presumably are close enough to Ann's claim. For some number of people facing paraplegia, we therefore ought to save them rather than Ann from death, since this maximizes the sum of strength-weighted, relevant claims.

Before pointing out the proposal's problems, I want to pause to emphasize that ARC, as I understand it, is a view about moral choice, and not about axiology. ARC will sometimes select outcomes that are worse in terms of overall goodness than other available outcomes.[9] ARC and Limited Aggregation in general, should only apply to the deontic assessment of choices and not to the evaluation of outcomes. As we will see later on, Limited Aggregation violates structural

[9] Voorhoeve himself writes that ARC will sometimes "have costs in terms of *impersonal* goodness." See Voorhoeve. p. 85, my emphasis.

properties that, in my view, are inherent to the very concept of goodness.[10] To keep things simple, I will assume the utilitarian axiology according to which the best outcome is the one maximising the overall unweighted sum of well-being. However, nothing hinges on this assumption, as long as we agree that axiology is a fully aggregative matter. If you prefer a prioritarian axiology, for example, then you can simply plug it in. Two further assumptions: First, I only consider choice under certainty, so each alternative act will be associated with exactly one outcome. And second, I only consider "same-people choices", i.e. all available outcomes contain the same set of people.

With these assumptions in place, we can now move on to the first set of objections against Limited Aggregation that I want to consider.

II. The Anchoring Ambiguity

The problem I deal with in this section is due to Patrick Tomlin.[11] It stems from an ambiguity in Voorhoeve's account that had not been appreciated before. The ambiguity is the following. There are (at least) two ways to understand the "strongest competing claim" (which Tomlin calls the "anchoring claim") to which any claim *c* must be close enough in order to be relevant. The first meaning is "the strongest claim with which *c* competes". The second meaning is "the strongest claim in the competition". These two meanings come apart when the

---

[10] Of course, some might disagree about this. See e.g. Temkin, *Rethinking the Good : Moral Ideals and the Nature of Practical Reasoning*.

[11] Patrick Tomlin, "On Limited Aggregation," *Philosophy & Public Affairs* 45, no. 3 (June 2017): 232–60.

two groups of claims that compete with each other are not homogenous in the sense that it is not the case that all claims within each group are exactly equally strong.[12] Take the following example.

> *Death and Fingers vs. Legs:* We can either save Ann from a terminal illness and prevent a number of people from losing a finger or prevent a number of different people from losing a leg.

What in this case determines the relevance of the claims of those that stand to lose a finger? Ann's claim or the claim of those that stand to lose a leg? Ann's "death-claim" (i.e. her claim to be saved from death) is the strongest claim in the competition, but this claim doesn't compete with the finger-claims. If we suppose that Ann's claim is the anchoring claim, then in Tomlin's terminology we "anchor by strength." The strongest claims with which the finger-claims actually compete, however, are the leg-claims. If we take those claims to be the anchoring claims, then we "anchor by competition." If we now plausibly assume that finger-claims are close enough to leg-claims, but not to death-claims, then they are irrelevant under the first interpretation, but relevant under the second.

Tomlin shows that both interpretations give rise to implausible results. First, consider Anchor by Competition and the following two-stage case.[13]

---

[12] I should note that Voorhoeve himself explicitly places such "diverse group cases" outside the scope of his original paper. However, it is clear that Voorhoeve doesn't think that ARC shouldn't apply beyond the "homogenous group cases" he restricts his discussion to and none of his arguments would support such a restriction.

[13] See Tomlin, "On Limited Aggregation." p. 240.

Two-Stage Case A:

Stage 1: We can save either Ann from a terminal illness (A) or n different people from losing a leg (B).

Stage 2: m people that we can save from losing a finger are added to each side.

I will illustrate this and following cases with simple figures. Each column represents one group we can save. The numbers (1, n, m in this case) represent the number of people holding a claim of a given strength. The higher up the number is in a given column, the stronger the claim. To keep things simple, I will assume that there are only three strengths a claim can take. There will only be death-claims, leg-claims, and finger-claims. If claims are represented in parenthesis, that means that they are not present at the first stage but are added at the second stage. *Two-Stage Case A* can then be represented as follows.

Two-Stage Case A:

A	**B**
1	
	n
(m)	(m)

Now suppose n is such that according to Limited Aggregation, at stage 1, we ought to save the people from losing a leg. Now we add m people that stand to lose a finger to Ann's side and to the side of those people that stand to lose a leg. As before, we assume that finger-claims are close enough to leg-claims, but not to death-claims. According to Anchor by Competition, this means that the claims of the people we added to Ann’s side are relevant, but the claims of the people we added to the other side are not. This seems counterintuitive, especially considering that at stage 2 we might have to reverse our judgement from stage 1 and should now save Ann from death and the *m* people on her side from losing a finger. This is because their relevant claims might turn the balance in Ann’s favour, while the irrelevant claims on the other side carry no counterbalancing weight. The approach violates what Tomlin calls

> *Equal Consideration for Equal Claims*. All claims of equal strength ought to be given equal weight in determining which group to save.[14]

Consider now Anchor by Strength by way of the following case, also adapted from Tomlin.[15]

> *Two-Stage Case B:*
>
> *Stage 1:* We can save either *n* people from losing a finger (A) or *m* people from losing a leg (B).
>
> *Stage 2:* Ann, whom we can save from suffering a terminal disease, is added to A.

A	**B**
(1)	
	m
n	

[14] Tomlin. p. 241.
[15] See Tomlin. p. 245.

Suppose again, that finger-claims are relevant to leg-claims, but not to death-claims and that *n* and *m* are such that at stage 1 we ought to save the people from losing a finger. It is obvious that adding Ann to this group ought to strengthen the case for saving them. Not only can we now save *n* people from losing a finger, but we can also save Ann's life and we obviously ought to do so. Anchor by Strength, however, might tell us otherwise. Since the strongest claims in the overall competition now no longer are the leg-claims, but Ann's death-claim, the finger-claims are no longer relevant and should play no role in our decision anymore. This means that we now have to weigh Ann's single claim against *m* leg-claims and depending on how large *m* is, ARC might tell us to satisfy those claims, which is clearly the wrong result. Anchor by Strength violates what Tomlin calls the

> *Principle of Addition*. Merely adding a claim to a group of claims cannot lessen that group's choiceworthiness compared with a fixed alternative.[16]

This principle seems intuitively plausible enough to reject any approach that violates it.[17] In his article, Tomlin goes over a number of possible, more complicated anchoring rules, which, however, suffer from even more devastating counterexamples.[18] This suggests that the

[16] Tomlin. p. 245.
[17] I should note, however, that there is a phenomenon, familiar from social choice theory, which would also violate the Principle of Addition. This is the so-called "No Show Paradox." Many otherwise attractive voting procedures have the implication that a voter could be better off "not showing up", rather than casting the ballot for her favourite candidate in the sense that her voting for her favourite candidate will actually lead to this candidate not being elected when she would have been elected had the voter abstained. See e.g. Hervé Moulin, "Condorcet's Principle Implies the No Show Paradox," *Journal of Economic Theory* 45, no. 1 (June 1988): 53–64.
[18] See Tomlin, "On Limited Aggregation." pp. 247-50.

approach of determining a set of relevant claims and then restricting aggregation to these claims is doomed to fail. In the following section, I therefore develop an alternative approach at fleshing out Limited Aggregation.

III. Aggregation with Constraints

I suggest the following specification of Limited Aggregation. In a choice situation with competing claims, we should look at the single strongest individual claim. If this claim is very much stronger than the strongest claim with which it competes, we do whatever satisfies this claim. If it is only slightly stronger, we do whatever is best in aggregate. Put differently, there is a general rule to do whatever is best in aggregate, i.e. realize the best outcome. However, the strongest individual claim can raise a veto. We then need to see whether this veto is justified or not. If it is justified, we do whatever satisfies this claim. If not, we proceed to realize the best outcome. Call this

> *Aggregation with Constraints.* If one person has much more at stake than any person whose interests conflict with hers, we ought to do whatever serves the interests of that person. Otherwise, we ought to do what realizes the best outcome.

This principle does better than ARC in Tomlin's cases. Reconsider

Two-Stage Case A:

A	B
1	
	n
(m)	(m)

On Aggregation with Constraints, in stage 1, Ann's claim isn't much stronger (given our assumptions) than any other person's competing claim, so we ought to do what realizes the best outcome. This depends on how large n is. The added finger-claims at stage 2, do not change this assessment. Aggregation with Constraints does not violate Equal Consideration for Equal Claims, at least not in this case. One case, in which one might think that it looks like it does violate the principle, is the following.[19]

Two-Stage Case C:

Stage 1: We can save either Ann from a terminal illness (A) or n different people from losing a finger (B).

Stage 2: m people that we can save from losing a leg are added to each side.

A	B
1	
(m)	(m)
	n

[19] I thank Theron Pummer for prompting me to discuss such a case.

At stage 1, Ann's claim is much stronger than any other claim, so by Aggregation with Constraints, we ought to save her. This judgement does not depend on how large *n* is. Now at stage 2 we may have to rethink our verdict. This is because the added leg-claim on the side of those threatened to lose a finger, competes with Ann's death-claim and, by assumption, is close enough to her claim. On Aggregation with Constraints we therefore now ought to do what leads to the best outcome. If *n* is large enough, we will now have to let Ann die and satisfy the finger-claims and the leg-claims we added to their side.

One might think that this violates Equal Consideration for Equal Claims. This is because, after all, we added the same number of leg-claims to each side and one could hold that since they carry equal weight, they should somehow "cancel each other out" and therefore together should do nothing to overturn our previous judgement at stage 1. The fact, however, that they do not cancel each out in this way, does not imply that we do not give the two claims equal consideration or even that we disregard the leg-claims we added to Ann's side. To see this, consider an instance of *Two-Stage Case C*, where *n* is such that on aggregation, the finger-claims together with the one leg-claim would just barely outweigh Ann's single death claim. The leg-claim we added to Ann's side now turns the tables in her favour. On aggregation it adds just as much force to her side as the other leg-claim added to the other side. It is only when *n* is large enough that saving these people from losing a finger would add more value to the outcome than saving Ann could, that we should overturn our judgement from stage 1. Only then is it the case that the added leg-claim to Ann's side can do nothing to help her. For all other *n,* it will turn out that we ought to help her and the person that was added to her side.

What is true, though, is that in cases like Two-Stage Case C, adding relevant claims to either side can "activate" previously irrelevant claims. It is like turning a switch that takes us from a non-aggregative to an aggregative mode. This implication may seem especially troublesome, when more claims are added to the group that will not be chosen than to the group that will be chosen. Take the following variation.

Two-Stage Case C:*

A	**B**
1	
(m)	(1)
	n

In this case, if n is very large, m could also be quite large, and we would still at stage 2 overturn our previous judgement. How problematic is this implication? As before, it does not violate Tomlin's Equal Consideration for Equal Claims. However, one might think that adding a claim to group B, while adding *more* claims of equal strength to an opposing group a, should not strengthen B's choiceworthiness compared to A.

I agree that at first blush this implication seems odd. However, on reflection, proponents of Limited Aggregation have a justification available. This is the thought that is at the very heart of Limited Aggregation. When we refuse to let a number of weaker claims outweigh one very strong claim, we do so because the person with the strong claim has so much more at stake than any person with a competing claim. Whenever one person (or a group of people) has much more at stake than any other person we could help instead, we ought to help this person irrespective of the numbers involved.

I take it that this is the driving intuition behind cases like *Death vs. Headaches* and it is fulfilled at the first stage, so we refuse the aggregation of the *n* finger-claims. The antecedent of this judgement no longer holds at the second stage, so by this judgement it is no longer true that we ought to help the person with the strongest claim, no matter what. There now is another person who has, by stipulation, a relevantly similar claim. We are going to help this person alongside many others with weaker claims instead, and so realize the best outcome possible.

It makes a difference whether a claim competes against relevantly similar claims or not. If it does not, then we refuse to aggregate. If it does, then what we ought do depends on the specific case and the specific version of Limited Aggregation. However, the fact that the existence of such a claim *could* make a difference to our mode of practical reasoning is not surprising.

The implication of Aggregation with Constraints in *Two-Stage Case C* is thus fundamentally different from the implication of ARC in Tomlin's *Two-Stage Case A*. In this case it was the addition of finger-claims that, by stipulation, are irrelevant to Ann's death-claim, which changed things. In *Two-Stage Case C* it is the addition of relevant leg-claims that changes things. This is a crucial difference, as we should expect relevant claims to have the power to change the normative situation we are in. We shouldn't expect irrelevant claims to have that power, however.[20] In essence, this is just a restatement of what it means to buy into Limited Aggregation. In certain situations, only some claims matter, others don't. Proponents of the view should therefore not be unnerved by this implication in *Two-Stage Case C.* Rather, this is what we should expect of a view of Limited Aggregation.

So far, I have shown that Aggregation with Constraints does not seem to run into the same problems as Anchor by Competition as it does not violate *Equal Consideration for Equal Claims.* To see whether it also does better than Anchor by Strength, consider now again Tomlin's second case,

Two-Stage Case B:

A	**B**
(1)	
	m
n	

At stage 1, there is no single claim that is much stronger than any other competing claim, so we ought to do what realizes the best outcome and this depends on n and m. If we now at stage 2 add Ann's death-claim to the finger claim, no matter how large n and m are, we won't have to now switch from satisfying the finger-claims to satisfying the leg-claims. If at stage 1 we ought to satisfy the finger claims, then at stage 2, we will still do so. Unlike Anchor by Strength, Aggregation

[20] For a similar line of thought see Victor Tadros, who proposes an otherwise very different view of Limited Aggregation. See Tadros, "Localized Restricted Aggregation."

with Constraints does not violate the Principle of Addition and therefore does better. Adding a claim to one group of claims without at the same time also adding claims to the competing group, can never lessen this group's choiceworthiness.

Here is the principled reason why my approach doesn't run into the problems raised by Tomlin. Whereas ARC aggregates some claims (the relevant ones) but not others (the irrelevant ones), my approach either aggregates all claims or none at all. It lacks the discriminating feature in ARC that gives rise to implausible results under both ways of resolving the ambiguity in ARC. My approach is thus not subject to the same objections as both ways of resolving the ambiguity in ARC.

There is, however, a cost to Aggregation with Constraints that some proponents of Limited Aggregation will likely be hesitant to accept. This is the fact that it gives a different verdict than ARC in a well-known case by Frances Kamm.

> *Sore Throat:* We can either save Ann from death or save Bob from death and cure Carl's sore throat.[21]

Within our framework, this case can be represented as follows.

A	**B**
1	1
	1

[21] See F. M. Kamm, *Morality, Mortality Volume I: Death and Whom to Save From It* (Oxford University Press, 1998). p. 146.

Kamm writes that the benefit of curing Carl's sore throat should be disregarded. It is an "irrelevant utility" and should play no role in our decision. Many proponents of Limited Aggregation agree.[22] According to Kamm's approach, we ought to treat this case just as we would treat the case were Carl not present. That is, we should be indifferent between saving Ann or Bob. On my approach, this is not so. Since Ann's claim and Bob's competing claim are equally strong, and therefore obviously close enough, we ought to aggregate and do what realizes the best outcome. This means that we should save Bob and cure Carl's throat, which realizes the best outcome.

While I did share Kamm's intuition when I first encountered her case, I am prepared to give it up on reflection. If we *really* are indifferent between two alternatives, then any additional reason that speaks in favour of one of the alternatives, but not the other, should make it the case that we are no longer indifferent. On Aggregation with Constraints, this is what happens when we add Carl's very small claim to Bob's side. Before we added Carl's claim, the reasons speaking in favour of saving Ann were *exactly* equally as strong as the reasons in favour of saving Bob. This no longer is the case once we added Carl. We now have more reason to save Bob and Carl than we have to save Ann and ought to do so.

As I said, this will come as a significant cost to some proponents of Limited Aggregation who are wedded to Kamm's intuition. Even so, this cost is, I think, much smaller than the costs of accepting the implications of ARC that Tomlin's cases brought out. So, Aggregation with Constraints still has the upper hand over ARC.

[22] See e.g. Voorhoeve, "How Should We Aggregate Competing Claims?"and Tadros, "Localized Restricted Aggregation."

Before I conclude my exposition of Aggregation with Constraints, however, there is one more case I want to discuss. Consider the following variation of *Sore Throat*.

> *Sore Throat 2:* We can either save Ann and Dave from death or save Bob from death and cure Carl's and *n* other people's sore throat.

Since in this case, it is not true that one person has much more at stake than any other person whose claim competes with hers, on Aggregation with Constraints, we should still save Bob and cure the sore throats if there are enough of them. This is the case although Dave, also facing death, is now on Ann's side. One could argue that unlike the original Sore Throat Case, Sore Throat 2 is a simple variant of Death vs. Headaches and therefore a stronger objection to Aggregation with Constraints.[23]

This case does pose a stronger challenge than the original Sore Throat Case. However, as I shall now show, it is similar to the challenge that Two Stage Case C posed and we can give essentially the same answer. We can see this by cashing out Sore Throat 2 as a two-stage case as well.

> *Two-Stage Sore Throat 2:*
>
> *Stage 1:* We can either save Ann from death (A) or cure *n* different people's sore throat (B).
>
> *Stage 2:* Dave, whom we can save from death, is added to Ann's side. Bob, whom we can also save from death, is added to the other side.

[23] I thank an anonymous referee of *Utilitas* for pressing me to discuss this case.

A	B
1 (1)	(1)
	n

At stage 1, it is clear that we ought to save Ann. This is just a reformulation of Death vs. Headaches (or, rather, Death vs. Sore Throats). What we ought to do at stage 2, however, now depends on how large n is. This is because, after adding Bob's death-claim, it is no longer the case that one person has much more at stake than any other person whose interests conflict with hers. Thus, on Aggregation with Constraints we now ought to do what is best in the aggregate. If n is large enough, we might have to overturn our judgement from stage 1.

This case has the same structure as Two Stage Case C, only that the added claims are of an even stronger kind, so by definition of the kind that is relevant in this situation. As I write above, these sorts of claims have the power to change the normative situation we are in. When we refuse to let a number of weaker claims outweigh one very strong claim, we do so because the person with the strong claim has much more at stake than any person with a competing claim. In this case, after we added Bob's claim opposing Ann's, this is no longer true. There is no one on either side who has much more at stake than any person with a competing claim. There are death-claims on both sides. This fundamentally changes the kind of justification we can give to the person whom we are not going to save. In a strictly pairwise comparison of the strongest claims, there would now be a tie where there wasn't one before. Therefore, the view that we should always satisfy the single strongest individual claim, which is the view that best embodies the non-aggregation sentiment, judges the two stages of Two-Stage Sore Throat 2 differently. Aggregation with Constraints does so as well.

This concludes my initial exposition of Aggregation with Constraints. We have seen that it does better with regard to some of the objections that have been levelled against other accounts of Limited Aggregation. There are, however, possible objections to Limited Aggregation that also apply to the view I proposed. I will take these up in the remaining two sections.

IV. Separability and Horton's Dilemma

Joe Horton has recently brought forward a new challenge to Limited Aggregation.[24] He proposes a dilemma that, in his view, is faced by all views endorsing Limited Aggregation. His challenge is important. This is because, both horns of his dilemma include a violation of a plausible separability condition, which he doesn't explicitly recognise. Separability says, roughly, that our choice between two distributions should not depend on elements that are the same in both distributions.

In this section, I show that Limited Aggregation does indeed violate this condition. This, however, is not actually surprising, but at the bedrock of the view's commitments. It brings to the surface the following: there is a tight connection between Limited Aggregation and relational factors as expressed by a violation of separability.

[24] Horton, "Always Aggregate."

This section proceeds as follows. First, I will introduce the notion of separability as it applies here and show that it is tightly connected to aggregation. Second, I explain how Horton's objection amounts to the claim that Limited Aggregation violates separability. Third, I argue that this violation is not actually an objection to, but rather a central feature of Limited Aggregation.

In general, separability is the idea that when evaluatively comparing two objects, we can ignore the objects' elements which are the same in both. It is a prima facie attractive principle and a powerful tool in ethics and decision theory. In decision theory, *separability of states of nature* is an important axiom of expected utility theory. It holds that when two alternative courses of action have the same outcome in some state of nature, then our choice should not depend on what happens in that state of nature.[25]

In ethics, we are more often concerned with the *separability of people*. This principle says that when comparing two distributions, we can ignore those people that fare the same in both distributions. In other words, our choice should be independent of the unaffected, where an individual *i* is unaffected just in case *i* has the same well-being level in both distributions. Separability is a practically convenient property, as it implies that when choosing between, say, two different policies, we can focus on those people actually affected by the policies. Here is a general formulation that serves our purpose.

[25] In decision theory, this is known as the "Sure Thing Principle". See e.g. Leonhard Savage, *The Foundations of Statistics*, 2nd ed. (Dover, 1972). See also chapter 1 of this thesis above.

> *Separability.* If two distributions have the same well-being level for one person, then our choice between them depends only on what they are like for other people.[26]

To illustrate the attractiveness of this principle, take the following case, where the numbers represent the well-being levels of three people, i_1, i_2, and i_3 and we have to choose between two distributions, D1 and D2.

	i_1	i_2	i_3
D1	2	2	1
D2	4	1	1

D1 vs. D2

Our choice between D1 and D2 might be affected by many considerations, for example by the fact that total well-being is greater in D2 than in D1. But what our choice should *not* be affected by, separability holds, is how well i_3 is doing. Her well-being is the same either way and should not affect our choice. This means that on Separability, D1 ought to be chosen rather than D2 if and only if (2,2) ought to be chosen rather than (4,1); and D2 ought to be chosen rather than D1 if and only if (4,1) ought to be chosen rather than (2,2).[27]

Separability plays an important role in some of the best positive arguments for utilitarianism and other fully aggregative theories, like prioritarianism.[28] As I show below, Horton's argument exploits that connection.

---

[26] For a more formal formulation and discussion, see Chapter 4 of Broome, *Weighing Goods*.
[27] I am indebted to an anonymous referee for *Utilitas* for pressing me to make this point and for improving the following discussion.
[28] For an argument for utilitarianism based on Separability, see e.g. Broome, *Weighing Goods*. For an argument for prioritarianism, see Matthew D. Adler, *Well-Being and Fair Distribution* (Oxford University Press, 2012). See also the

However, before applying Separability to the kinds of cases we are dealing with here, we need to take care of a slight complication. Since i_3 would not have a claim by the light of our theory, she would not occur in the choices we have been considering thus far. In these choices we only considered the well-being of people who have claims on us. And since, by definition, only people whose well-being we can affect, can have claims, i_3 would not be considered. We thus need a slightly stronger principle to show how Limited Aggregation violates Separability. We need a principle that tells us that neither i_3 nor i_4 should affect our choice in a case like the following.

	i_1	i_2	i_3	i_4
D3	2	2	1	0
D4	4	1	0	1

D3 vs. D4

Such a principle is a little harder to come by, as both i_3 and i_4 have a stake in our decision. i_3 has a claim to D3 and i_4 has a claim to D4. So in some sense, we should not ignore them. However, it is clear that in the following case, where the decision only includes i_3 and i_4, we should be indifferent between D5 and D6.

	i_3	i_4
D5	1	0
D6	0	1

D5 vs. D6

We have no reason to favour i_3 over i_4 or vice versa and both have exactly the same stake in our decision, so it is clear that we should be

arguments I outline in chapter 1 above, which rely on certain "independence principles" very similar to Separabilty as outlined here.

indifferent between both options.[29] We might then think that attaching such indifferences to our choices, should not change anything. In other words, when choosing between two distributions, our decision is unaffected by those elements between which we are indifferent. Call this

> *Separability*.* If two distributions have subdistributions that we are indifferent between, then our choice between them depends only on what they are like in other subdistributions.

This principle implies that the presence of i_3 and i_4 should not affect our choice between D3 and D4 above. Prima facie, this principle will be attractive to many people. Like other separability principles, however, it is violated by some theories in distributive ethics, most notably egalitarian theories.[30] This is because, for these theories, it is important how individuals fare *relative to others*. It is also violated by Limited Aggregation and for much the same reason.

For example, to see that Aggregation with Constraints violates Separability*, we can just take another look at

---

[29] This is also given by the following principle.
Weak Anonymity. If two outcomes *A* and *B* differ only in that the identities of two people have been permuted, we should be indifferent between *A* and *B*.
I don't appeal to this general principle here, as it is likely to be rejected by many aggregation sceptics. However, even these sceptics will accept the indifference in this restricted case.

[30] In fact, a violation of separability is what distinguishes "true" egalitarian theories from other distribution sensitive theories. See John Broome, "Equality versus Priority: A Useful Distinction," *Economics and Philosophy* 31, no. 02 (July 2015). 219–28. I agree with Broome on this.

Two-Stage Case C:

A	B
1	
(m)	(m)
	n

A and B are equally choiceworthy in the respect that they both include the same number of people with leg-claims, m. Between these two subdistributions we are indifferent. According to Separability* this should then not affect our choice. However, we have seen that under Aggregation with Constraints, adding these claims *does* affect what we should do.

Now, Horton's dilemma is the following. The first horn: He shows that views of Limited Aggregation either have implications like my Aggregation with Constraints has in Two-Stage Case C above. So, a violation of Separability* is what gives rise to the first horn of Horton's dilemma. The second horn is the following: We can construct cases where, according to Limited Aggregation, we ought to save group A over B, and group C over D, but in a choice between the union of groups A and C and the union of groups B and D, we ought to choose B and D.[31] In other words, we are violating a principle we can call

> *Horton's Principle.* If we ought to choose distribution A over B and distribution C over D, then we ought to choose distribution (A&C) over (B&D).[32]

[31] See Horton, "Always Aggregate." 11.

[32] Horton does not explicitly mention this principle, but it is clear from his discussion that this is what he has in mind.

The specifics of Horton's case need not concern us here, but he is correct that views of Limited Aggregation can violate this principle. Underlying this implication, however, is, again, a violation of separability. This is not recognized by Horton.

Horton's principle follows from Separability* and transitivity. If we ought to choose A over B, then by Separability* we ought to choose (A&C) over (B&C). And if we ought to choose C over D, then by Separability* we ought to choose (B&C) over (B&D). If the transitivity of "ought to be chosen over" holds in this case,[33] it follows that we ought to choose (A&C) over (B&D), which is Horton's principle.[34] So, what gives rise to both horns of Horton's dilemma is the same thing: a violation of a separability. This means that Horton's objection amounts to the claim that all views of Limited Aggregation violate separability.

How surprising is this violation though? To see that it is not very surprising and actually a central part of the view, we just need to look at some of the different formulations of Limited Aggregation. According to Voorhoeve's ARC, "[a] claim is relevant if and only if it is sufficiently strong *relative to* the strongest competing claim."[35] And according to my Aggregation with Constraints, "if one person *has much more at stake than any person whose interests conflict with hers*, we ought to do whatever serves the interests of that person." On views of Limited Aggregation, what we should do, depends in part on how people fare relative to others. That these views then violate

---

[33] I say "in this case", because I don't want to assume that the transitivity of this relation holds universally. See the next section.
[34] I am indebted to Kacper Kowalzcyk for discussion of this point in particular and the issues in this section more general.
[35] Voorhoeve, "How Should We Aggregate Competing Claims?": p. 66. My emphasis.

Separability is not a surprise or an unwelcome implication, but rather a central building block of the theory. Limited Aggregation is a *relational view*.

It is important to re-emphasize, however, that this does not imply anything about the underlying axiological structure. Aggregation with Constraints is a deontic view and the separability violations I have outlined occur only at that level. In fact, I am sceptical that any plausible ethical view could have this structure at the level of axiology.[36]

This concludes my discussion of Horton's objection to Limited Aggregation. It wasn't my aim to invalidate the objection. Rather I wanted to show that it is not so much an objection to Limited Aggregation as a restatement of a central part of the view. I argue elsewhere in more detail that there is a tight connection between this violation of Separability and scepticism about aggregation. There I argue that aggregation sceptics should be egalitarians, thus violating separability.[37] For the remainder of this article, however, I now want to turn to a different set of objections that I need to address before Aggregation with Constraints can stand as a contender for the most plausible version of Limited Aggregation.

V. Cases with three or more Alternatives

This objection is brought out when we apply my principle not to a binary choice but to a choice between three (or more) alternatives.

[36] I tend to agree with Broome, who argues that, in his terminology, "good is coherent". See e.g. Chapter 6 of Broome, *Weighing Goods*.
[37] See chapter 1 of this thesis, "Aggregation and Equality."

Given the apparently intransitive structure of Limited Aggregation and the violation of plausible axioms of rational choice that follow from this structure, it is not surprising that Limited Aggregation faces problems when extended beyond binary choices. For Limited Aggregation, the relation "ought to be chosen over" is not always transitive. After all, as we have seen in many examples, it is the very idea of Limited Aggregation that, if the numbers are right, we ought to save n people from losing an arm rather than Ann from death, and m people from losing a finger rather than n people from losing an arm, *but we should not* save m people from losing a finger rather than Ann from death. This structure violates transitivity and in some cases in turn gives rise to a violation of other plausible principles of rational choice. Here, I do not want to add to these well-rehearsed general criticisms and responses to them.[38] Rather I want to focus on one case of this kind which seems to pose a challenge specifically for Aggregation with Constraints.

> *Death and Finger vs. Death vs. Fingers:* We can either save Ann from a terminal illness and Bob from losing a finger, Carl from the same terminal illness, or *n* people from losing a finger. [39]

[38] For some of the criticisms see e.g. Parfit, "Justifiability to Each Person.", John Halstead, "The Numbers Always Count," *Ethics* 126, no. 3 (April 2016): 789–802. Alastair Norcross, "Intransitivity and the Person-Affecting Principle," *Philosophy and Phenomenological Research* 59, no. 3 (1999): 769–76. For responses see F. M. Kamm, *Intricate Ethics* (Oxford University Press, 2007). pp. 297-98, 484-87, Voorhoeve, "How Should We Aggregate Competing Claims?", pp. 76-79, Alex Voorhoeve, "Why One Should Count Only Claims with Which One Can Sympathize," *Public Health Ethics*, March 3, 2016, 1–9. Tomlin, "On Limited Aggregation." fn. 11.

[39] Variations of this case were presented to me by Theron Pummer and Bastian Steuwer.

A	B	C
1	1	
1		n

In this case, Aggregation with Constraints seems to imply that we ought to choose C and save the people from losing a finger if there are enough of them. This is because, unlike in, for example, *Death vs. Headaches,* it is no longer the case that "one person has much more at stake than any person whose interests conflict with hers" and so according to Aggregation with Constraints we ought to realize the best outcome, which by assumption means to satisfy the finger-claims, if n is large enough. One could plausibly argue that this is just as implausible as the claim that there is some n such that we ought to satisfy the finger-claims when the only other option is to save Ann, which would be equivalent to giving up our guiding intuition from *Death vs. Headaches*.

Another problem with a case like this is the following. If we consider pairwise comparisons of the three alternatives, we get a clear winner: A. Aggregation with Constraints would clearly choose A over C. It would also choose A over B. It is odd that the same view would then choose C out of all three options.

As I write above, because of their structural properties, such problems are bound to arise for any view of Limited Aggregation. However, if we are unsettled by this particular case, we could amend Aggregation with Constraints in a way that takes care of this case and that is in line with the rationale of the view.

This is to first eliminate all alternatives that do not include a claim that is close enough to the strongest claim. For *Death and Finger vs. Death vs. Fingers* this would mean to eliminate C, and choose A out of the remaining two alternatives. That is, we exclude all options that could not be chosen in a pairwise comparison with the alternative that includes the strongest claim(s), irrespective of the numbers involved. Only then we apply Aggregation with Constraints as outlined above. This procedure is in line with the central idea that the strongest claim overall has special significance. The justification of our choice needs first and foremost be directed at the person who has the greatest stakes in our decision.

Conclusion

In this essay I have clarified the notion of Limited Aggregation and have shown its limitations. I have shown where its most prominent account, Vorhooeve's Aggregate Relevant Claims, fails and have proposed an account that does better. I have called this Aggregation with Constraints. My proposal is a much more modest version of Limited Aggregation. If we are attracted to the initial idea, I take it this is the best we can hope for. However, the proposal manages to retain the central thought at the heart of Limited Aggregation. This is the thought that sometimes very strong individual claims stand in the way of doing what would be best in the aggregate and Ann gets to live in *Death vs. Headaches*. Of course, my view stands and falls with the idea that whether a person has much more at stake than any other person is very important. It thus puts even more pressure on the intuition in *Death vs. Headaches*. But we knew that all along. If this intuition weren't so robust, there would be no point in trying to come up with a theory of Limited Aggregation.

3 | On Ex Ante Contractualism*

Contractualism is a claims-based model of moral rightness. It is the view, brought forward most notably by T.M. Scanlon, that an action is right if and only if it is justifiable to all. An action is justifiable to all just when it is licensed by a principle that cannot be reasonably rejected by any single individual.[1] Further, a principle can only be reasonably rejected for personal reasons. Contractualism thus construed excludes impersonal reasons derived from, for example, the overall value of an outcome. It thereby denies the permissibility of interpersonal aggregation of harms and benefits to determine which action is right. In situations where individuals have competing claims to be helped, we always ought to pursue the policy that satisfies the single strongest individual claim, or, in converse, minimizes the strongest individual complaint against it, by following "the principle whose implications are most acceptable to the person to whom it is least acceptable."[2]

This implication of contractualism clearly demarcates the view from aggregative theories like utilitarianism. I here understand

---

* An earlier version of this chapter has been published as Korbinian Rüger, "On Ex Ante Contractualism," *Journal of Ethics and Social Philosophy* 13, no. 3 (2018).

[1] See Scanlon, *What We Owe to Each Other*. pp. 189-248.

[2] Rahul Kumar, "Risking and Wronging," *Philosophy & Public Affairs* 43, no. 1 (2015): 27–51. p. 31.

utilitarianism as standard act utilitarianism, where we always ought to pursue the action that will lead to the greatest (expected) sum of well-being. The difference between the two rival theories becomes apparent in cases like

> *Death vs. Headaches:* We can either save Ann from a terminal illness or prevent some number of different people from suffering a mild headache.

By virtue of what Ann stands to lose, her claim to be saved from death is clearly greater than any other individual claim to be spared a headache. Under contractualism, we therefore ought to save her. This is the case irrespective of how many people stand to suffer a headache. Under utilitarianism, on the other hand, our answer will depend on the number of people that we could spare the headache. For some number of people, the benefits derived from the spared headache will *in sum* outweigh the benefit to Ann if we choose to save her. Contractualism demands what many people take to be the obviously correct choice in *Death vs. Headache*.[3]

This is straightforward in hypothetical situations of absolute certainty like the above. The approach, however, is less clear about situations in which we don't yet know the outcomes our choices will lead to. These cases, however, are much more common. With Barbara Fried, one could even say that

> in the real world, no conduct, judged ex ante, is certain to harm others. This is true even of harms that are intended […]. If I point a gun at your head and pull the trigger, I am overwhelmingly

[3] For empirical data along these lines, see e.g. Voorhoeve, "Why One Should Count Only Claims with Which One Can Sympathize."

likely to kill or seriously injure you, but I am not certain to do so. The gun could misfire, I could have forgotten to load it [etc.].[4]

So rather than occurring with certainty, most harms result from risks that have been imposed on people or have not been eliminated. It is therefore imperative for contractualists to offer an account of how their theory deals with risk.

Johann Frick has developed such an account: *ex ante contractualism*.[5] In brief, ex ante contractualism holds that in situations involving risk we ought to act in accordance with principles that license the action that satisfies the strongest individual claim, where those claims are a function of the expected value that a given policy gives each person ex ante. It thus offers an alternative to the ex post reasoning employed by other contractualists, most notably Scanlon himself.[6]

I here challenge Frick's version of ex ante contractualism on contractualist grounds. My argument proceeds as follows. In the first section, I distinguish between ex ante and ex post contractualism in more detail. In the second section, I argue that adopting ex ante contractualism would have far reaching implications that contractualists would find very hard to accept. I show that ex ante

---

[4] Barbara H. Fried, "Can Contractualism Save Us from Aggregation?," *The Journal of Ethics* 16, no. 1 (2012): 39–66. p. 50.

[5] Frick, "Contractualism and Social Risk." See also Johann Frick, "Uncertainty and Justifiability to Each Person," in *Inequalities in Health*, ed. Nir Eyal et al. (Oxford University Press, 2013), 129–46. Unless noted otherwise, henceforth "ex ante contractualism" refers to Frick's version of the view.

[6] See Scanlon, *What We Owe to Each Other*. pp. 189-248. See also Sophia Reibetanz, "Contractualism and Aggregation," *Ethics* 108, no. 2 (January 1998): 296–311. Note that Scanlon has since changed his position, crediting an earlier version of Frick's article. See T.M. Scanlon, "Reply to Zofia Stemplowska," *Journal of Moral Philosophy* 10 (2013): 508–514. For a critique of the ex post approach see Elizabeth Ashford, "The Demandingness of Scanlon's Contractualism," *Ethics* 113, no. 2 (2003): 273–302.

contractualism in fact includes an implicit appeal to the interpersonal aggregation of harms and benefits. In the third section, I show that Frick's argument for the principled priority of identified over unidentified lives, another troubling implication of ex ante contractualism, is unsound. In the fourth and final section, I briefly comment on a possible pluralistic approach to get around some of the defects of ex ante contractualism. I conclude that to deal with uncertainty, contractualists should not adopt ex ante contractualism, at least not in Frick's version. Rather, they should adopt a suitably amended ex post approach.

I. Ex Ante and Ex Post Contractualism

Let me introduce the ex post approach and then contrast it with the ex ante approach by way of one of Frick's examples.

> *Mass Vaccination*: One million children are threatened by a virus, which will kill all of them if we do nothing. We must choose between producing one of three vaccines:
>
> Vaccine 1 is certain to save every child's life. However, if a child receives Vaccine 1, the virus will permanently paralyze one of the child's legs.
>
> Vaccine 2 gives every child a 99.9% chance of surviving the virus completely unharmed. However, for every child there is a corresponding 0.1% chance that Vaccine 2 will be completely ineffective. (Assume that the outcomes for different children are

probabilistically independent.) Call the children who end up dying the *luckless children*.

Vaccine 3 is sure to allow 999,000 children to survive the virus completely unharmed. However, because of a known particularity in their genotype, Vaccine 3 is certain to be completely ineffective for 1,000 identified *doomed children*.[7]

First consider a choice between only vaccines 1 and 3 (V1 and V3). Here we are not dealing with uncertainty and it is straightforward what contractualism recommends. If we choose V1, no single child will have a complaint that is as strong as the individual complaints of the doomed children if we choose V3.[8] We therefore ought to choose V1. If on the other hand, we consider a choice between V1 and V2, we are entering the territory of risk and things are less clear. This is because there are two ways of singling out the relevant complaints that we should take into account. Under one interpretation, we look at the outcome that a given vaccine will produce and look at the single strongest complaint any individual will have in that outcome. If we

---

[7] See Frick, "Contractualism and Social Risk." pp. 181-83. Note that Frick presents two distinct cases, in both of which Vaccine 1 is available, but Vaccines 2 and 3 only are available in one case.

[8] Frick writes: "The individual burden of becoming paralyzed in one leg, though significant, is not even close to that of losing one's life at a young age." (Frick. p. 183) Note that this information underspecifies (or even ill-specifies) the strength of the individual complaints. If we assume a counterfactual account of harm, the complaints of the doomed children if we pick V3 over V1 are not complaints against being left to die, where the alternative would be life in full health, but complaints against being left to die, where the alternative would be life with one paralysed leg. Such complaints are presumably much weaker. I think this point is overlooked by Frick. Nonetheless it is reasonable to assume that these weaker complaints are still decisively stronger than the complaints of the other (non-doomed) children against a policy that leaves them with one paralysed leg, where the alternative would be life in full health.

choose V2, we expect 1000 children to die.[9] Though we do not know how many children exactly will die, it is statistically certain that at least one child will die.[10] Since we are concerned with the single strongest individual complaint, this is all we need to know. Like in V3, this complaint will be stronger than any complaint under V1. Again, we ought to choose V1. This is the ex post approach.

According to the ex ante approach on the other hand, the relevant complaints are a function of the expected value an action gives each individual before it is performed. Under this account, a complaint against being subjected to a risk of suffering a harm is the complaint against being subjected to that harm with certainty discounted by the unlikelihood of the harm actually occurring. In *Mass Vaccination* the individual ex ante complaints against V2 are thus only 0.1% as strong as a complaint against dying from the virus with certainty. The strongest ex ante complaint against V2 is therefore much smaller than the strongest ex ante complaint against V1, which in turn is smaller than the strongest ex ante complaint against V3.

Accordingly, Frick's account selects V2 over V1, V1 over V3, and V2 over V3. This ensures that in each choice we minimize the strongest ex ante complaint. The ex post approach on the other hand would choose V1 over V2 and V3, and would likely be indifferent between V2 and V3.

Mass Vaccination thus shows how the ex ante and ex post approaches come apart. According to Frick, it also shows why ex post contractualism is unattractive. It fails to make a principled distinction

[9] $0.001 \times 1{,}000{,}000$.

[10] $1 - \left(\frac{999}{1000}\right)^{1000000}$.

between V2 and V3. As long as we know that *someone* will die if we pick V2, and therefore has a stronger complaint than anyone if we had chosen V1, this is enough for ex post contractualism to rule out V2. It fails to take into account the special predicament the doomed children find themselves in under V3 as it assimilates their fate to those of the luckless children in V2. Frick would say, it fails to distinguish between the fact that "we know that someone will die" (V2) and the fact that "there is someone whom we know will die" (V3).[11]

Because of these alleged shortcomings of ex post contractualism, Frick proposes his ex ante approach. The main argument for this approach is the "argument from the single person case."[12] According to this argument, if we have an option available that is in the best ex ante interest of all individuals, we ought to choose it. We ought to adhere to the following principle:

> *Weak Ex Ante Pareto*. If prospect P has higher expected value than prospect Q for all *i*, then we ought to choose P over Q.

---

[11] See Frick, "Contractualism and Social Risk." p. 200 and Johann Frick, "Treatment versus Prevention in the Fight against HIV/AIDS and the Problem of Identified versus Statistical Lives," in *Identified versus Statistical Lives: An Interdisciplinary Perspective*, ed. I. Glenn Cohen, Norman Daniels, and Nir Eyal (Oxford University Press, 2015). p. 193. I don't take this distinction to be morally as important as Frick's thinks it is. I shall not argue for this claim directly, though. Rather I will show that there are cases where even Frick's own account fails to make the distinction.

[12] See Frick, "Contractualism and Social Risk." pp. 186-94. See also Frick, "Treatment versus Prevention in the Fight against HIV/AIDS and the Problem of Identified versus Statistical Lives." p. 133. For similar arguments also see Dougherty, "Aggregation, Benificence and Chance." and Caspar Hare, "Should We Wish Well to All?," *Philosophical Review* 125, no. 4 (October 2016): 451–72.

Frick argues that the argument from the single person case establishes this principle as a principle of contractualist ethics. We can decompose cases like *Mass Vaccination* into a large number of single-person gambles. Suppose again that we are facing a choice between V1 and V2 (recall that ex post contractualism chooses V1). This choice can be broken down into one million single-person cases. Suppose that Ann is one of the affected children and we ask ourselves, what we would choose if we were solely motivated by her self-interest. We know that V1 will let her survive the virus but leave her with one paralysed leg, and that V2 will let her survive the virus completely unharmed with probability 99.9% and will lead to her death with probability 0.1%. Given reasonable assumptions about which level of well-being (or "utility") these three possible outcomes would deliver, we can calculate the expected value of both options. Suppose, we assume that for Ann life with one paralysed leg is 4/5 as good as life at full health, which we can arbitrarily fix to well-being level 10, with death corresponding to 0. The expected value of V1 then is 8,[13] while the expected value of V2 is 9.99.[14] Thus, the expected value of V2 for Ann exceeds that of V1 and if we are only concerned with her best interest, we ought to choose V2.[15] This seems to be the right course of action. After all, what other than Ann's best interest would we base our decision on?

But, of course, this reasoning is correct for every single child in *Mass Vaccination*, where the possible outcomes and corresponding odds are exactly the same as in the one-person case. Thus, if we are concerned

---

13 $\frac{4}{5} \times 10.$

14 $\frac{999}{1000} \times 10 + \frac{1}{1000} \times 0.$

15 Note that this result will be achieved even if Ann considers life with one paralysed leg only slightly worse than life at full health.

with every child's best interest, we ought to choose V2, just like we ought to choose V2 in the one-person case, when we are only concerned with Ann's best interest. The contractualist rationale behind this is that choosing V2 is the only action that is justifiable to all. Whatever the outcome of choosing V2, we can offer each child the following justification: "When we had to choose, we did what was in your own best interest." This justification is not available to us if we choose V1. I confess that I find this argument very seductive. In the following I argue, however, that contractualists ought to reject it and with it Weak Ex Ante Pareto.

II. The Implications of Ex Ante Contractualism

Return to *Mass Vaccination*. Only now suppose that instead of V3, we have V3* available. Like V3, V3* is sure to allow 999,000 children to survive the virus completely unharmed. However, because of a certain particularity in their genotype, V3* is certain to be completely ineffective for 1,000 *unidentified* doomed children, instead of *identified* doomed children. We can imagine, for example, that we have tested all 1,000,000 children for that genotype and have found out that the vaccine will be ineffective for exactly 1,000 of them. However, before we communicated the test results to anyone, our system broke down and we now have no way of assigning the positive results to any particular children.

Given that we chose V2 over V3, should we now choose V2 or V3*? In order to answer that question, we need to investigate whether V3* is relevantly different from V3. Only if it is, can we justify choosing V2 over V3, but being indifferent between V2 and V3*. If it is not then, given that we chose V2 over V3, we also ought to choose V2 over V3*. In this section I will argue that first, we should not judge V3 and V3* differently, second, that ex ante contractualism, however, is committed to doing so and, third, that this puts the account in a precarious position.

To me V3* seems like V3 in all important respects. In V3* as in V3, we know the exact outcome. We know that exactly 1,000 children are going to die and that for them the vaccine was always going to be ineffective. Like with V3, these children are doomed to die if we choose V3*. I therefore fail to see why we should choose V2 over V3, but be indifferent between V2 and V3*. This amounts to the claim that we should prefer to have V3* rather than V3 available. To see that this is implausible, suppose that we not only have V3* but two different vaccines V3*, and V3** available. However, for each of these vaccines it will (very likely) be an entirely different group of 1,000 children for whom the vaccine will not work and who will be killed by the virus. Obviously we have no reason to choose any one of these vaccines over the other. Whatever we do, 1,000 unknown, doomed children are going to die. The vaccines are equally choiceworthy and we should randomize.

Suppose that we settle on V3*. Before we actually administer the vaccine, however, we learn who the children are for whom V3* will not do anything (maybe we were able to restore our data base for V3*). Should we now because of that switch to V3**? I think clearly not. This would be an unnecessary "second lottery" and would arbitrarily favour those children for whom V3* is ineffective to the disadvantage of those children for whom V3** is ineffective. Nothing about the vaccines has changed and we said above that they are equally choiceworthy. They still are. This, however, is in effect the same situation we face when comparing Frick's V3 and my V3*. We therefore ought not to judge V3 and V3* differently.

Frick's account, however, is committed to judging V3 and V3* differently. It is committed to judging V3 impermissible, but V3* (along with V2) permissible. This is because the argument from the single person case applies to V3* as it applies to V2. Here too, it would be in each individual child's best interest to choose the risky vaccine over V1. The expected value of V3* for each individual child is the same as V2's. [16]

Though Frick doesn't consider V3*, he considers a nearby case. This case is like my V3*, only that there is a test we could carry out to identify the doomed children. The test, however, would be very expensive. Frick argues that in this case, administering the vaccine would be justifiable to all and therefore permissible. In such a case, we can say to each child "given justifiable limits on the resources we can be expected to expend in gathering further information about your particular case, [the vaccine] is highly likely to benefit you, and has

---

[16] Ex ante contractualists could reply that there is one important difference between V2 and V3* that I have overlooked. Namely that while in V3* it is merely *epistemically* uncertain who will die, in V2 it is *objectively* (or *physically*) uncertain who will die. Frick, however, carries out his discussion on the assumption that *all* probabilities are merely epistemic. He writes that "[w]hen using the terms 'probability' or 'chance' […] I assume that we are speaking not about objective indeterminacy at the level of physical reality itself, but about epistemic probability." (Frick, "Contractualism and Social Risk."p. 182). He furthermore argues, rightly I think, that for the moral assessment of risky policies this distinction makes no difference (Frick. pp. 197-201). In any case, it is doubtful whether objective probabilities at the physical level even exist (see e.g. David Lewis, "A Subjectivist's Guide to Objective Chance," in *Studies in Inductive Logic And*, ed. Richard Jeffrey (University of America Press, 1980), 267–297. Letting one's moral theory depend on the assumption that they do exist seriously diminishes its attractiveness.

only a tiny chance of turning out to your disadvantage."[17] If administering the vaccine when it is *very costly* to find out which children will not be helped by it, is permissible, then a fortiori administering V3*, when it is *impossible* to find out which children carry the problematic gene, must also be permissible.

I now argue that this judgement concerning V3* spells trouble for the account. Consider the following case.

> *Glass Box Villain (known victim):* An evil villain has taken 26 hostages named Ann, Bob, Carl … and Zeta. He places you in the following diabolic choice situation: he has placed all of them in 26 individual glass boxes stood up side by side. The last box is made out of regular glass and the other 25 boxes are made out of extra heavy glass. You can see that Zeta is placed in the last box. The villain asks you to decide between the following two options:
>
> (1) he will either fire a shot at her box, or
>
> (2) fire 25 individual shots at the other boxes. If he fires at Zeta's box, the bullet won't be stopped and Zeta will be killed.
>
> If he fires at the 25 boxes made of extra heavy glass, the glass will divert the bullets. However, the glass will crack and the debris will disfigure the 25 hostages in a way that permanently leaves them at a well-being level 9.5 on a scale from 0 to 10, where 10 corresponds to a life in full health and 0 to death. If you refuse to decide, the villain will blow

[17] Frick, "Contractualism and Social Risk." p. 194.

> up all boxes, killing all 26 hostages. How should you decide?

I assume that refusing to decide should be ruled out as an option. Between the two remaining options, it is clear what contractualism tells you to do. You should choose (2). Choosing (1) would kill Zeta only to save 25 other people from a relatively minor harm. The complaints of the 25 on you are not even close to Zeta's complaint. Since contractualism prohibits you from aggregating the 25 weak complaints to outweigh Zeta's strong complaint, you ought to save Zeta's life and let the villain fire at the 25 boxes made of extra heavy glass. Since there is no uncertainty involved, ex ante and ex post contractualism don't come apart in this case. Consider, however the following variation of the case.

> *Glass Box Villain (unknown victim):* Everything is as before, only now the boxes are opaque and neither you nor the hostages know whether it is Zeta or any of the other 25 in the box made of regular glass. How should you decide?

I think if in *Glass Box Villain (known victim)* you ought to stop the villain from firing at the last box, then in this case you ought to act in the same way. I cannot possibly see why the fact that the 26 boxes are now opaque should change our moral assessment of the case in any way. (Remember that there is nothing about Zeta as a person that should make us favour *her* over the other 25 hostages in any way.) However, Frick's ex ante contractualism is committed to the view that while in *Glass Box Villain (known victim)* you ought save Zeta, in *Glass Box Villain (unknown victim)* you ought to let the villain kill the person in the last box.

It is so committed because in *Glass Box Villain (unknown victim)* for all you know it could be any of the 26 hostages in the last box. In this respect, it is parallel to V3* above. You have no reason to assume that any of the 26 was more likely to end up there than anyone else. As far as you know, for each of them, there is a 1/26 chance that they are the one in the last box and a corresponding 25/26 chance that they are among the ones in the boxes made of extra heavy glass. This means that for each of them if you let the villain fire at the last box, there is a 1/26 chance that they will die and a 25/26 chance that they walk away completely unharmed. If you choose otherwise on the other hand, for each hostage there is a 1/26 chance that they walk away unharmed (if they are the one in the last box) and a 25/26 chance that they walk away slightly, but permanently disfigured.

I have arbitrarily assumed that this disfigurement leaves them at level 9.5 out of 10.[18] If we also assume that death leaves the hostages at "level" 0, then the choice situation can be represented by the following table

	S1		**S2**		**S3**		**…**	**S26**	
	A	**Others**	**B**	**Others**	**C**	**Others**	**…**	**Z**	**Others**
First 25 boxes	10	9.5	10	9.5	10	9.5	…	10	9.5
Last box	0	10	0	10	0	10		0	10

You have two available actions (again, ignoring the option of doing nothing): "first 25 boxes" and "last box." There are 26 equiprobable states of the world (S1-S26), corresponding to the 26 possibilities of who could be the one in the last box, where S1 corresponds to the state

[18] If you think that this is too low or too high, then you can adjust the level and change the number of hostages accordingly without affecting the basic structure of the case.

of the world where Ann is the one in the last box, S2 to the state where Bob is the one, and so on. The table shows the well-being levels of the hostages for each of these 26 states and the two available actions. For example, if you decide on "first 25 boxes" and Ann is the one in the last box, then she will be left at level 10, corresponding to full health, while the other 25 hostages (the "others") will be left at level 9.5. On the other hand, if you decide on "last box" and Ann is the one in that box, she will die ("level 0") and the others will be left unharmed at level 10. We can now calculate the expected value for each hostage under each of the two available actions. If you choose "first 25 boxes", the expected value is 9.52.[19] If you choose "last box", the expected value is 9.62.[20] Since 9.62 is greater than 9.52, if you want to do what's in each of the hostages' best interest, you ought to choose "last box." If you could ask them, they would want you to do so, or if for each of the hostages there was a guardian present who is only motivated by their beloved interest, they would tell you to do so! Therefore, via the argument from the single person case, ex ante contractualists (and proponents of Ex Ante Pareto in general) are committed to letting the villain kill the person in the last box. Note that here ex ante contractualism is so committed although "there is someone whom we know will die." We know that there is a person we will willingly sacrifice, namely the person in the last box.

This fact points to an objection that could be pressed against my exposition: it is not in fact true that every hostage has a 1/26 chance of being the one in the last box.[21] At the time of decision there is a fact of

[19] $\frac{1}{26} \times 10 + \frac{25}{26} \times 9.5.$

[20] $\frac{25}{26} \times 10.$

[21] This was suggested to me by Jeff McMahan and Tom Sinclair.

the matter who the person in that box is. From this it follows that it is not actually true that choosing "first 25 boxes" is in the best interest of everyone.

I think this objection will not succeed, at least not for an ex ante contractualist of Frick's kind. This is because this same objection could be pressed against someone, like Frick, who distinguishes between V3 and V3* above, deeming V3 impermissible and V3* permissible. In both V3* and V3, there is a fact of the matter who the children are that are going to die. The only difference is that in V3* informational constraints keep us from knowing the identities of these children. The same holds for *Glass Box Villain (unknown victim)*. So, if one thinks that we should not distinguish between *Glass Box Villain (unknown victim)* and *Glass Box Villain (known victim)* because in both cases there is a fact of the matter who is in the last box, then we also ought not to distinguish between V3 and V3*, because here in both cases there is also a fact of the matter who the 1,000 children are for whom the vaccine will do nothing.[22] As things stand, ex ante contractualists are committed to choose "first 25 boxes" in *Glass Box Villain (known victim)* and "last box" in *Glass Box Villain (unknown victim).*

I think this result should worry ex ante contractualists, especially since qua contractualists, they would be deeply committed to choose otherwise in *Glass Box Villain (known victim)*. The case thus lays bare the implications of the view that its proponents need to accept. These are implications many contractualists, or many non-consequentialists in general for that matter, find hard to stomach. *Glass Box Villain (unknown victim)* shows that ex ante contractualists need in some cases

[22] This, of course, is the position I am arguing for. It is however not available to ex ante contractualists, as they *want* to distinguish between V3 and V3*.

be prepared to sacrifice a person's life in order to protect many other people from a relatively minor ailment. This strikes me exactly as the kind of interpersonal aggregation that contractualism set out to avoid in the first place.

Now, ex ante contractualism's proponents might be prepared to bite the bullet. They could say that the fact that the number of people affects each individual prospect (holding everything else fixed) is simply directly implied by the way ex ante contractualism is defined. Frick calls this "counting the numbers without aggregating."[23] One could thus object to my exposition that I am implicitly assuming what I intend to show, namely that ex ante contractualism cannot be correct. For if one instead assumes that ex ante prospects are what we should be concerned with in a case like *Glass Box Villain*, then it plainly follows that we should order the villain to fire at the last box. To some extent this objection is warranted, for I *am* assuming that a theory that tells us let the person in the last box be killed in *Glass Box Villain (unknown victim)* should strike contractualists as dubious, if not wrong. The point is that rather than embracing this "number counting" as a welcome implication of the view, contractualists should be worried about a view that has these implications since it allows the numbers of people on each side of a binary choice to affect what we ought to do, even though the individual benefits and burdens are not affected.

The reason why most contractualists (and other nonconsequentialists) are opposed to interpersonal aggregation is because it violates what following Rawls has become to be called the "separateness of

[23] Frick, "Contractualism and Social Risk." p. 201.

persons."[24] According to one very strict version of this thesis, the aggregation of harms across different individuals is meaningless since there is no single entity to suffer the aggregate harm. As C.S. Lewis writes,

> [s]uppose that I have a toothache of intensity x: and suppose that you, who are seated beside me, also begin to have a toothache of intensity x. You may, if you choose, say that the total amount of pain in the room is now 2X. But you must remember that no one is suffering 2X: search all time and space and you will not find that composite pain in anyone's consciousness. There is no such thing as a sum of suffering, for no one suffers it.[25]

This, however, is exactly what ex ante contractualists overlook in *Glass Box Villain (unknown victim)*. If we let the number of people in the boxes made of extra heavy glass affect what we believe we ought to do, then we are overlooking the fact that the harms that any of the hostages is going to suffer doesn't increase or decrease with that number.

Might ex ante contractualists respond to my argument so far by claiming that there is a principled difference in importance between saving an identified person and saving an unidentified person that I have overlooked? If so, this difference could explain why we should in fact let the villain kill the person in the last box in *Glass Box Villain (unknown victim)*, while we should stop him from killing Zeta in *Glass Box Villain (known victim)*, as well as explain why we should choose V2 over V3, but be indifferent between V2 and V3*. In the following section I investigate this possibility.

---

[24] See John Rawls, *A Theory of Justice (Revised Edition)* (Oxford University Press, 1999). p. 167.

[25] C.S. Lewis, *The Problem of Pain* (Collins/Fontana Books, 1957). pp. 103-104.

III. The "Pro-Identified Lives Argument"

Many people attach greater importance to saving identified lives than to saving unidentified lives.[26] It is doubtful, however, that this psychological fact is of any moral relevance.[27] I, for one, do not think it is. It will have to be, however, in order to justify ex ante contractualism's way of distinguishing between V3 and V3* as well as between *Glass Box Villain (unknown victim)* and *Glass Box Villain (known victim).* Luckily for ex ante contractualists, Johann Frick offers an ingenious argument to that effect. He argues that correctly applying the ex ante contractualist rationale to cases that are "competitive ex ante" yields the conclusion that we ought to prioritize identified over unidentified lives. In this section I attempt to show that this argument does not succeed.

In section 1, we have seen how ex ante contractualism coincides with the ex ante pareto principle in cases where there are actions that are in the ex ante interest of everyone. The principle, however, does not apply in cases that are competitive ex ante. Here every action that is in the interest of one group of people comes at a cost to another group of

---

[26] See R Moore, "Caring for Identified versus Statistical Lives: An Evolutionary View of Medical Distributive Justice," *Ethology and Sociobiology* 17, no. 6 (1996): 379–401. and George Loewenstein and Karen Jenni, "Explaining the 'Identifiable Victim Effect,'" *Journal of Risk and Uncertainty* 14 (1997): 235–257.

[27] See T. Schelling, "The Life You Save May Be Your Own," in *Problems in Public Expenditure Analysis*, ed. S.B. Chase (Brookings Institution Washington D.C., 1968), 127–62. D. W. Brock and D. Wikler, "Ethical Challenges In Long-Term Funding For HIV/AIDS," *Health Affairs* 28, no. 6 (November 1, 2009): 1666–76. and Michael Otsuka, "Risking Life and Limb: How to Discount Harms by Their Improbability," in *Identified versus Statistical Lives: An Interdisciplinary Perspective*, ed. I. Glenn Cohen, Norman Daniels, and Nir Eyal (Oxford University Press, 2015).

people even at the ex ante stage. Take the following example employed by Frick.[28]

> *Miners*: A single miner, Jones, is trapped in a mineshaft and if we don't help him, he will die. The rescue mission, however, would be very costly. These resources could instead be used to make the mine safer for everyone working there in the future. Suppose there are 100 other people working at the mine and with the resources we would have to use on the rescue mission, we know that we could instead reduce their risk of suffering a fatal accident from 3% to 1%. What should we do?[29]

If we decide to let Jones die and make the mine safer for future workers, we can expect to save two workers' lives in the future[30] instead of saving Jones' life now. Frick argues that ex post contractualists in this case are committed to letting Jones die and saving the two other workers' lives instead.[31] This is because, no matter what we do, the strongest individual complaints are equally strong in both cases. These are the complaints of the miners that will die when we could have prevented it, Jones in the one case and the unnamed two miners in the other case. And since under Scanlon's contractualism "numbers break ties" when the strongest complaints are equally strong on both sides, we ought to do what satisfies the greater number of strongest claims.[32]

---

[28] See Frick, "Contractualism and Social Risk." p. 212.

[29] Further assume that we know that no one else but these 100 people will ever work at the mine.

[30] 100×0.02.

[31] See Frick, "Contractualism and Social Risk." p. 214.

[32] Scanlon's "tie breaking argument", where he draws on an argument by Frances Kamm (see Kamm, *Morality, Mortality Volume I*. pp. 101, 114-19 and Scanlon, *What We Owe to Each Other*. pp. 229-41) is contested. (See e.g. Michael Otsuka, "Scanlon and the Claims of the Many versus the One," *Analysis* 60, no. 3 (2000): 288–93.) However, despite the defects of this particular argument, I find it highly plausible that when deciding between one claim on the one

Again, Frick thinks ex post contractualism goes wrong here. He offers his "pro identified lives argument" to show why this is so and takes this argument to provide a principled defense of the claim that we ought to prioritize identified lives over unidentified lives. The argument starts from the premise that, in general, people have a stronger claim to be saved from suffering a harm with certainty than to be saved from suffering that same harm with some probability $p<1$.[33] This claim is undoubtedly correct. Suppose we have to decide between saving Ann from certain death or reducing Bob's risk of death from 3% to 1%. It is clear that we ought to help Ann in this case. Now Bob's claim to have his death risk reduced is identical, Frick continues, to each of the 100 miners' claims in *Miners*. From this it follows, that no individual miner has a stronger complaint than Jones. Coupled with the contractualist ban on interpersonal aggregation, it follows that we ought to minimize the single strongest complaint and save Jones.

Frick claims that, first, this argument provides a principled defense for the privileging of identified over unidentified lives and, second, that it also shows where ex post contractualism goes wrong. He claims that ex post contractualists are committed to the view that in *Miners* there is someone that has a stronger claim than Bob in the 1 vs. 1 case. He writes, "[s]omehow, the fact that, if we save [Jones], it is foreseeable that *someone* from the group of 100 will die in a future accident is thought to strengthen the complaint of whoever turns out to be

---

hand and two claims of equal magnitude on the other hand, we ought to satisfy the two claims.

[33] See Frick, "Contractualism and Social Risk."p. 215. See also Frick, "Treatment versus Prevention in the Fight against HIV/AIDS and the Problem of Identified versus Statistical Lives."pp. 188-191.

harmed."[34] This, Frick argues, is an implicit appeal to interpersonal aggregation over "different possible worlds."

Regarding the first point: I think that the argument does not provide a principled defense for the favouring of identified lives in general, but only in a very narrow class of cases like *Miners*. It only provides a defense for favouring an identified person *when and because* that person holds a claim that is stronger than any competing claim. It, for example, does not provide a defense of the type needed to justify the ex ante contractualist's choices in *Glass Box Villain (unknown victim).* Here, the dialectic of comparing ex ante claims and then satisfying the single strongest claim does not work, since here all ex ante claims are equally strong, as we have seen. It thus fails to provide a justification for why it is more important to save Zeta in *Glass Box Villain (known victim)*, than to save the unidentified person in the last box in *Glass Box Villain (unknown victim).* This is because the argument does not provide a principled defense for the claim that it is more important to save an identified person rather than and unidentified person *because* that person is identified. Such a defense, however, would be needed to justify ex ante contractualism's verdicts in the *Glass Box Villain* cases.

Regarding the second point: First of all, it is not clear that ex post contractualism really is committed to letting Jones die in *Miners*. We don't *know* that two miners will die in the future if we decide to save Jones. Yes, this is the expected outcome, but, of course, it is only one of many different possible outcomes. The chance that *exactly* two miners will die is only around 27%[35] and the chance that *at least* two miners

[34] Frick, "Contractualism and Social Risk." pp. 217.
[35] $\binom{100}{2} \times 0.02^2 \times 0.98^{98}$.

die, so as to tip the scale in favour of letting Jones die, is around 60%.[36] No part of ex post contractualism commits proponents of the view to disregard these probabilities entirely. Frick assumes that they would take the expected outcome of an action and then simply act as if they knew that that expected outcome would actually eventuate. This, of course, would be a mistake. By doing so they would not be able to differentiate between cases like *Miners* and a case where we have to decide between saving one person from certain death and saving two different people from certain death. But I don't think that anything commits them to this precarious position. Instead, they could take into account the likelihood of enough miners dying so as to outweigh Jones' claim. I take this to be the most plausible interpretation of ex post contractualism.[37] As we have seen, the likelihood of at least two miners dying is only 60%. So why should we just assume that ex post contractualists wouldn't rescue Jones?

Second of all, as I have argued before, the main problem with many instances of interpersonal aggregation of harms is that any sum of weaker harms together does not constitute anything meaningful, since there is no one suffering from this aggregate harm. This, however, is not the case in *Miners*. In this case, the aggregate of the many trivial harms *is* suffered by a single individual. The more people work at the mine, the likelier it becomes that *someone* will die as a result of us not making the mine safe. This is a different kind of aggregation. Contrast this with a variation of *Miners*, where we can either save Jones or use the resources to distribute life-long supplies of Aspirin to all future miners, who occasionally suffer headaches because of the stuffy air in

[36] Pr (100 deaths) - Pr (0 or 1 deaths).
[37] See chapter 4 of this thesis below.

the mine. This aggregation is more like the kind of aggregation employed by the ex ante contractualist in *Glass Box Villain (unknown victim).* In this case, as we have seen, the number of people involved has no effect on the harm that the most burdened individual has to suffer. As long as we lack an independent objection against this second, different kind of aggregation, I don't see why ex post contractualists need to be moved by this particular argument.

IV. Pluralism as a Way out?

Let me now turn to the final problem with ex ante contractualism that I want to raise in this essay. This problem is acknowledged by Frick himself. Consider a variation of *Miners*, only now there are 1,000 other miners in addition to Jones. Call this *Miners 1,000*. In this case, we would expect 20 miners to die in the future if we save Jones now. Frick submits that ex ante contractualism here "goes too far."[38] For him, it is clear that, given some number of expected deaths (which could be greater or lesser than 20), we ought to let Jones die. Frick concedes that this problem for ex ante contractualism can only be solved "by scaling back the ambitions of contractualism as a moral theory."[39] He argues that in cases where his theory is unable to yield the intuitively correct verdicts, it should be assisted by other non-contractualist principles. As a candidate, Frick suggests that we should take into account the effect an action has on people's well-being in general. For example, in

[38] See Frick, "Contractualism and Social Risk." p. 219
[39] Frick. p. 219

Miners 1000 we should take into account that there will be "a much greater loss of life" if we save Jones.[40]

This sounds like Frick is suggesting that the contractualist should call utilitarianism to her rescue when her theory fails her intuitions. This ad hoc move, however, is available to ex post contractualists as well. They, too, can be pluralists about interpersonal morality. Like Frick, they, too, can say that in some cases their theory needs to be assisted by impersonal concerns to decide what the right course of action is. I see no reason why, prima facie, it should seem more plausible to restrict ex ante contractualism in such a way than it is to restrict ex post contractualism in the same manner. The only difference being that it would be different cases that the theory can deal with "on its own."

In *Mass Vaccination*, for example, facing a choice between V1 and V2, ex post contractualists can say the following: "In principle we ought to choose V1 here, since this minimizes the largest complaint ex post. However, the consequences of doing so in terms of overall well-being are too grave to be ignored. After all, if we do choose V1, we will leave 1,000,000 children with only one functioning leg for the rest of their lives. This overall loss in well-being is much greater than if 1,000 children die prematurely." If ex ante contractualists can legitimately resort to these impersonal reasons when their theory yields intuitively unattractive implications, the same route should be open for ex post contractualists as well.

However, I have misgivings about this ready resort to pluralism. As Frick himself notes, Scanlon's theory itself is already pluralist in a

---

[40] See Frick. p. 222.

way.[41] He limits his contractualism to the domain of "what we owe to each other." We might call this domain of morality "interpersonal morality" or, following Frances Kamm, "M1."[42] However, Scanlon deems his theory to exhaust this part of morality. Frick, on the other hand, thinks that contractualism should be assisted by other principles even within this already limited domain. The question then is how valuable contractualism is as a theory above and beyond these other principles. I suspect that it is no longer very valuable. Rather, it seems unacceptably ad hoc and gerrymandered to fit a very narrow class of cases. Whenever we look beyond this narrow class of cases and the theory fails to yield the right result, its proponent can resort to pluralism. Leaving the theory open in this way, however, limits its value. It means that the theory has too many free parameters, limiting its predictive power and testability, thereby putting into doubt its value as a standalone moral theory.

Moreover, even granting this pluralistic approach, it will not get ex ante contractualism around the implications of the *Glass Box Villain* case of the second section. Utilitarianism here pulls in the same direction as ex ante contractualism. All other things equal, the overall aggregate value of an outcome where 25 people are spared a 0.5 unit decrease in well-being is higher than the value of an outcome where one different person is spared a 10 unit decrease in well-being. So, even if the misgivings I have with the pluralistic approach are unwarranted, this problem remains.

[41] See Frick. fn. 47 on p. 220.
[42] Kamm, *Intricate Ethics*. pp. 455-90.

Conclusion

In this essay, I have challenged Johann Frick's ex ante contractualism. I argued that adopting the view leads to implications contractualists will find hard to stomach. This has become especially vivid in *Glass Box Villain (unknown victim)*, where ex ante contractualists are committed to sacrificing one person in order to save 25 different people from relatively minor harm. I have argued that this is an instance of the kind of interpersonal aggregation of harms, contractualists sought to avoid in the first place. I also argued that this kind of aggregation is more troublesome than the kind of aggregation Frick accuses ex post contractualists of. In connection to this last point I have argued that Frick's argument for the principled priority of identified over unidentified lives also fails, because it can only account for ex ante contractualism's verdict in a very narrow class of case. Finally, I have argued that Frick's resort to pluralism is ad hoc and further unable to block some of the unwelcome implications of the view. I conclude that if there is no other way of developing ex ante contractualism that does not run into these problems, contractualists ought to be concerned with the probability that harm could befall someone, rather than with the probability that harm could befall a specific person. For contractualists, a suitably amended ex post approach is better equipped to honour this commitment.

4 | Minimizing the Strongest Complaint under Risk

According to the most common extensional specification of contractualism, in situations in which individuals have competing claims to our aid, we always ought to pursue the policy that satisfies the single strongest individual claim or, equivalently, minimizes the strongest individual complaint. There has recently been much discussion about how this and other non-aggregative models should be extended to cover cases in which we face empirical uncertainty.[1] I here spell out what I take to be the most plausible extension to conditions of risk.[2]

In the literature, we can find broadly three alternatives regarding how contractualists could go about extending their theory to conditions of risk.[3] First, they could take an ex ante approach. According to this

[1] See e.g. Reibetanz, "Contractualism and Aggregation." Frick, "Contractualism and Social Risk." Kumar, "Risking and Wronging." Joe Horton, "Aggregation, Complaints, and Risk," *Philosophy & Public Affairs* 45, no. 1 (2017): 54–81. Joe Horton, "Aggregation, Risk, and Reductio," *Ethics*, forthcoming. Alec Walen, "Risks and Weak Aggregation - Why Different Models of Risk Suit Different Types Cases," *Ethics*, forthcoming.

[2] Under these conditions, though we are uncertain what will happen, we can at least assign probabilities to all states of nature. For the purposes of this essay, I take these probabilities to be merely epistemic.

[3] Note that nothing of what I say in in the following is, in principle, restricted to contractualism. In the recent literature on aggregation and risk, the term "contractualism" has, somewhat unfortunately, become known as a catch-all label for *any* theory that gives the same verdict as Scanlonian contractualism

approach, individual complaints are a function of the expected value that a given policy gives each person ex ante and we ought to minimize the strongest ex ante complaint.[4] Second, they could take an ex post approach. We then take the relevant complaints to be a function of the actual results of our policy. That is, the strongest complaint against a given policy is raised by the individual actually burdened most by it.[5] Third, they could go for a theory that seeks to reconcile the ex ante and the ex post approaches, by arguing that some cases call for the ex ante approach and other cases call for the ex post approach.[6]

I have previously argued against Johann Frick's contractualist theory, which is the best developed ex ante approach.[7] While this essay does add to my earlier critique against the ex ante approach, I largely focus on the latter two alternatives. I argue that the hybrid approach is neither theoretically well motivated nor able to combine the strengths rather than the weaknesses of both "pure" approaches. This leaves the ex post approach, which stands on motivationally firmer ground. The approach, however, is itself subject to a number of important objections. I introduce the ex post approach in some detail and reply to these objections.

---

in certain contentious cases involving numbers, i.e. is a non-aggregative theory.

[4] The foremost proponent of this approach is Johann Frick. See Frick, "Contractualism and Social Risk." See also Frick, "Uncertainty and Justifiability to Each Person."

[5] For an approach along those lines, see Otsuka, "Risking Life and Limb: How to Discount Harms by Their Improbability."

[6] See Walen, "Risks and Weak Aggregation - Why Different Models of Risk Suit Different Types Cases."

[7] See Frick, "Contractualism and Social Risk." For my critique, see Rüger, "On Ex Ante Contractualism."

The essay proceeds as follows. In the first section, I introduce different kinds of cases involving risk, as they appear in the recent literature. The intuitive judgements we make in these cases are often seen as lending support to one over another approach to moral decision making under conditions of risk. We can distinguish between what I call *same risk* and *different risk* cases, as well as *known distribution* and *unknown distribution* cases. This gives us four categories of cases: *same risk, known distribution*; *same risk, unknown distribution*; *different risk, known distribution*, and *different risk, unknown distribution*. I show that the distinctions between these categories of cases are either not as clear-cut as they first appear or else not morally relevant. This casts doubt on proposals according to which different approaches should govern different kinds of choices. A recent proposal of this kind by Alec Walen will be introduced and rejected for this reason in the second section.

In the third section, I introduce the ex post approach in more detail. I start with what I call the *naïve ex post approach*. This approach disregards uncertainty entirely and is obviously implausible. A more plausible approach has been proposed by Sophia Reibetanz. This account, while a clear improvement on the naïve approach, turns out to be nothing else than a hybrid approach of the kind I criticised above. We should therefore reject it for the same reasons. I then proceed to a more sophisticated account along the lines of an approach by Michael Otsuka. According to this approach, we ought to discount complaints by the probability that *someone* will be in a position to raise them against our choice. This allows us to take into account complaints that are uncertain to arise. However, I show that this account suffers from two acute problems. I propose an account that does not run into the same problems. In the fourth and final section, I defend this account against two objections.

I. A Taxonomy of Cases

The recent literature on the ethics of risk features a great wealth of different stylized cases. Johann Frick's "Vaccination Cases", Joe Horton's "Villain Cases", and a series of cases by Michael Otsuka have been particularly influential.[8] The sheer number of different cases and the fact that the cases often only differ in tangential features, makes it hard to keep track of the relevant features that distinguish them. In this section, I show that we can organize these cases along two dimensions. One is the *same risk/different risk* distinction, the other one is the *known distribution/unknown distribution* distinction.

Start with the following case, which is structurally similar to Johann Frick's by now well-known *Mass Vaccination (Unknown Victims)*.[9]

[8] See Frick, "Contractualism and Social Risk." Horton, "Aggregation, Complaints, and Risk." Horton, "Aggregation, Risk, and Reductio." and Otsuka, "Risking Life and Limb: How to Discount Harms by Their Improbability."
[9] See Frick, "Contractualism and Social Risk." pp. 181-82. One difference from Frick's case is that this case involves risk in both options.

Same Risk, Unknown Distribution: The well-being of 1,010,000 people is at stake.[10] We have to choose between α and β.

(α) All are subjected to an independent 99% risk of losing a leg.

(β) All are subjected to an independent 0.1% risk of death.

This case is symmetric in the sense that all people are in the same position ex ante. They all face the *same risk*. Whatever we can say about one person's predicament, applies to all. It is an *unknown distribution* case in the following sense. We can imagine, for example, that (β) amounts to the following lottery. Each individual draws a ticket from an urn containing one red ticket, corresponding to death, and 999 black tickets, corresponding to life. They then put their ticket back before the next person draws theirs. Such a lottery implies that even anonymously speaking, we are not sure about what the outcome will look like. If we opt for (α), it is our *expectation* that around 1,000,000 people would lose one leg.[11] But, of course, we are unsure about how many people exactly would lose one leg. If we opt for (β) on the other hand, we expect around 1,000 people to die.[12] However, again, we are unsure about how many people will actually end up dying. This is because the probabilities involved are independent. This implies that,

---

[10] Note that for reasons of representational simplicity, in describing this case, as well as the following cases, I do without telling an elaborate backstory as to why and how the well-being of these people are at stake. As long as we assume that they are all innocent, equally deserving and so on, we need no viruses or villains to see the structural features of a given case.

[11] More exactly, the number of "expected lost legs" is 999,900 or $0.99 \times 1{,}010{,}000$.

[12] Again, more exactly, the number of "expected deaths" is 1,010 or $0.01 \times 1{,}010{,}000$.

in fact, it is very unlikely that *exactly* 1,000 of them would die, and there is even a tiny chance that no one would die. However, this chance is so small that it is a statistical certainty that at least one person will die, which is all we really need to know.

This last fact implies that according to the ex post approach, we ought to choose (α), as this would minimize the largest complaint ex post. A complaint against death is stronger than a complaint against losing a leg. According to ex ante contractualism, on the other hand, we ought to choose (β), as this would minimize the strongest ex ante complaint. For we can reasonably assume that a complaint against losing a leg with 99% probability is stronger than a complaint against being subjected to a 0.1% risk of death. Moreover, given that the case is same risk in the above sense, we can apply the following principle.

> *Weak Ex Ante Pareto*. If prospect P has higher expected value than prospect Q for all *i*, then we ought to choose P over Q.

In the above case, (β) has higher expected value for every person than (α) and should thus be chosen, according to this principle. So, the ex post approach violates Weak Ex Ante Pareto.

Weak Ex Ante Pareto is intuitively attractive. In effect, it says that if everyone affected would rationally want us to act in a certain way, then we should do so. Acting against Weak Ex Ante Pareto would thus require overriding the presumed consent of the people on whose behalf we are choosing. As Matthew Adler and Chris Sanchirico put it, "in some choice situations, the ex post planner will stand alone against the unanimous wishes of her constituency."[13] This, of course, is a large bullet to bite. Weak Ex Ante Pareto, however, is violated not just by the ex post approach, but by many other theories as well, most notably ex post egalitarian and prioritarian theories.[14] I shall have more to say about Weak Ex Ante Pareto and the ex post approach's violation of it further below in section four. For now, let us move on to a different category of risky case. Consider this case by Joe Horton. [15]

> *Villain 2:* A villain has kidnapped C1 and ten billion other people. We must choose between the following.
>
> (α) He will cut one hand off C1.
>
> (β) He will randomly select and kill ten of the other people.

This case is different from *Same Risk, Unknown Distribution* and Frick's case in two ways. First, it is not the case that all people are in the same position ex ante. C1 is in a very different position from the ten billion other people. In fact, there is a clear conflict between what is in C1's best interest ex ante and what is in everyone else's best interest ex ante. It is what I will call a *different risk* case.

Second, it is not an unknown distribution case in the above sense. Rather, it is what I will call a *known distribution* case. The risks faced by the ten billion people are dependent on each other and negatively

---

[13] Matthew D Adler and Chris William Sanchirico, "Inequality and Uncertainty: Theory and Legal Applications," *University of Pennsylvania Law Review* 155, no. 2 (2006): 279–377. p. 347.

[14] For an in-depth analysis see chapter 7 of Adler, *Well-Being and Fair Distribution*. See also chapter 1 of this thesis, "Aggregation and Equality".

[15] See Horton, "Aggregation, Complaints, and Risk." p. 58. I have slightly rephrased the case for representational reasons. This case can be seen as an antipole to Frick's Vaccination case and thus my *Same Risk, Unknown Distribution*, as it elicits opposite intuitions.

correlated. If we choose (β), then we subject ten billion people to a one-in-a-billion risk of being killed. The risk involved, however, is different from an unknown distribution case. We could imagine, for example, that in this particular case, the urn contains ten billion tickets, each corresponding to one person. The villain draws exactly ten tickets. In this case then, we are certain about what the outcome will look like, anonymously speaking, if we choose (β). C1 will keep his hand, but ten other people will lose their lives. The only uncertainty involved is with regard to the question *who* these ten people will be. Horton's *Villain 2* is thus a *different risk, known distribution* case.

It is another case in which the ex ante and ex post approaches conflict. The ex ante approach would here opt for (β).[16] C1's complaint against having his hand chopped off, is clearly stronger than any of the ten billion complaints against a one-in-a-billion risk of death. This means that we minimize the strongest complaint ex ante by choosing (β). The ex post approach, on the other hand, tells us to choose (α). We know that choosing (β) will give rise to ten ex post complaints against death, each of which is stronger than C1's complaint. So, we minimize the strongest ex post complaint by choosing (α).

[16] I should note that Frick's own approach avoids this and similar implications by "scaling back the ambitions of contractualism as a moral theory," as Frick puts it. See Frick, "Contractualism and Social Risk." p. 219. As I pointed out in chapter 3 of this thesis, this move is, of course, also available to the ex post approach to avoid its implications in a case like *Same Risk, Unknown Distribution*. More on this below in section 4. For now, I focus on extensions of contractualism to conditions of risk that retain its core commitments and ambitions as a non-aggregative moral theory that tells us what is right or wrong *all things considered*.

From Horton's case we can see that though the ex ante approach respects Weak Ex Ante Pareto, it violates a different plausible principle.

> *Weak Stochastic Dominance*. If we ought to choose every possible outcome delivered by prospect P over every possible outcome delivered by prospect Q, then we ought to choose P over Q.[17]

Though a little cumbersome, the attractiveness of this principle is easy to grasp. A bit roughly, it says that if, whatever happens, you know that you will prefer to have chosen A over B, then you should choose A over B. Or, put in a slogan, when you know that you would regret choosing a particular alternative, then you should not choose that alternative. This is a very modest principle. However, the ex ante approach violates it. A case like Horton's *Villain 2* shows that.

In this case, if we choose (β), there is an astronomically large number of equiprobable ways the world could go.[18] This is the number of possible combinations of selecting ten out of the ten billion people. In each of these possible states of the world, ten particular people will be killed. We also know what would happen in each state of the world if

---

[17] I borrow the name of this principle from Matthew Adler, who employs a parallel axiological principle. See Adler, *Well-Being and Fair Distribution*. p. 495. This is the weakest possible of a number of conceivable such dominance principles. Another common, slightly stronger, principle is *Statewise Dominance*. This principle says, roughly, that if in every possible state of the world, A delivers an outcome that ought to be chosen over the outcome delivered by B *in that same state*, then A ought to be chosen over B. Alex Voorhoeve and Marc Fleurbaey employ a similar principle to argue against the ex ante pareto. They call it the "Principle of Full Information." See Marc Fleurbaey and Alex Voorhoeve, "Decide As You Would with Full Information!," in *Inequalities in Health*, ed. Nir Eyal et al. (Oxford University Press, 2013), 113–28.

[18] $\binom{10{,}000{,}000{,}000}{10}$.

we choose (α): C1 loses his hand. This means that whichever way the world goes, in isolation, the outcome produced by (α) would clearly be chosen by contractualists over every outcome that (β) could have produced. Weak Stochastic Dominance says that this implies that we ought to choose (α) over (β), even if we do not yet know which state of the world actually materializes.

To me, this principle is overwhelmingly plausible. Along with even stronger versions of it, it is also widely accepted among decision theorists and economists.[19] I assume that violating this principle is unquestionably a big bullet to bite for any theory. What is arguably most problematic about theories which violate this and related principles, is that they are prone to time inconsistencies.[20]

Suppose you employ the ex ante approach and you are in a situation in which it instructs you to choose B, which is stochastically dominated by A. This means that by the theory's lights, you ought to choose B, even though you know that *whichever way* the uncertainty gets resolved, by the same theory's lights you will wish you had chosen A instead. This means that the theory is crucially dependent on *when* decisions are taken. At time t_0, the theory tells you to choose B, and at time t_1, after the uncertainty is resolved, it would tell you to choose A. Since you know all this, it seems irrational to still choose B at t_0. As

---

[19] See Chris Starmer, "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature* 38, no. 2 (June 2000): 332–82. p. 335.

[20] See e.g. Adler and Sanchirico, "Inequality and Uncertainty: Theory and Legal Applications." at 344-347. Larry G. Epstein and Uzi Segal, "Quadratic Social Welfare Functions," *Journal of Political Economy* 100, no. 4 (August 1992): 691–712. and Marc Fleurbaey, "Welfare Economics, Risk and Uncertainty," *Canadian Journal of Economics/Revue Canadienne d'économique* 51, no. 1 (February 2018): 5–40.

John Broome writes about theories exhibiting this kind of structure, by choosing B at t_0, you

> choose rightly, but it later turns out you chose wrongly. Indeed, it may turn out that you ought later to undo what you rightly did. Moreover, you might be able to foresee as you choose [B] that just this would happen. This is a most implausible sort of incoherence in your activity.[21]

Another way to see the irrationality of this is the following. After having chosen B at t_0 and the uncertainty is resolved, you might at t_1 be willing to incur a cost, for example in form of an additional harm, in order to switch to the outcome that A would have brought about. However, you could have had that same outcome without the extra cost by choosing A at t_0. Knowing all of this and still choosing B at t_0 is implausibly irrational.

One way in which proponents of the ex ante approach could respond to this is by saying that if choosing a given option was the right thing to do by the lights of the ex ante approach, then you shouldn't regret having chosen the right option even once the identities of the victims are revealed. Neither should you want to reverse your choice. It remains the right thing to have done. It is a mistake to assume that the uncertainty at the time of choosing is a normatively irrelevant feature of this choice situation.[22]

This response suggests that proponents of the ex ante approach should take something like a *resolute* approach to diachronic choice. Roughly, according to this approach, decision makers should make a plan at t_0

---

[21] John Broome, *Weighing Lives* (Oxford University Press, 2004). p. 75.
[22] Johann Frick raised this point to me in conversation. He defends this line of reasoning in more depth in Frick, "Uncertainty and Justifiability to Each Person." This article responds to Fleurbaey and Voorhoeve, "Decide As You Would with Full Information!"

as given to them by their theory and then stick to it. They should stick to it even if, at some later point in time, their theory would tell them to choose differently if they were given the chance. As Edward McClennen writes,

> suppose that [the resolute chooser] proceeds, against the background of his decision to adopt a particular plan, to do what the plan calls upon him to do, even though it is true (and he knows it to be true) that were he not committed to choosing in accordance with that plan, he would now be disposed to do something quite distinct from what the plan calls upon him to do. I propose to describe such an agent as a resolute chooser.[23]

A resolute proponent of the ex ante approach could say that, given the fact that, by the lights of my theory, uncertainty about *who* will be harmed or benefitted affects the expected value of a given action and thus affects the strongest ex ante complaint, it is morally significant. This is taken into account when adopting a plan that the resolute decision maker then sticks to, even if, at a later stage, he would have to try to overturn this plan, taking that later stage in isolation.

Now, while it seems clear that uncertainty at the time of choosing *can* be a normatively relevant feature, it is much less clear that such uncertainty is normatively relevant in cases in which we can "peek ahead" and see that the uncertainty will be resolved in such a way that one option leads to an unequivocally preferable state of affairs than another available option. I propose that on the ex ante approach, the uncertainty about who will be harmed and who will be benefitted in cases like *Villain 2* is of this latter kind. In these cases, a resolute approach seems much less attractive.

---

[23] Edward F. McClennen, *Rationality and Dynamic Choice: Foundational Explorations*, 1st ed. (Cambridge University Press, 1990). p. 13.

Even Frick, as a proponent of the ex ante approach, agrees with this in some cases. These are cases in which we can adopt an "ex ante rule" that at that time is everyone's interest, but later would license the agent to act in a way that seems clearly impermissible. Here is Frick's own example.

> *Human Experiment*: At [t_0], a doctor organizes a lottery among a group of 100 paraplegic children, which selects 10 of them by a random process. At [t_1], he conducts medical experiments on these 10 children, foreseeing (but not intending) that they will die in the course of the experiments. He knows for certain that the knowledge gained in this way will allow him to cure the remaining 90 children of their paraplegia.[24]

Frick claims that going through with the lottery and the experiment is impermissible. I agree. Still, adopting the plan to do so would be in everyone's best interest at t_0, and thus licensed by the ex ante approach. A resolute decision maker guided by the ex ante approach would adopt the plan at t_0 and then stick with it. Frick wants to avoid this implication and proposes the following:

> *Decomposition Test*: If a rule or procedure can be decomposed into a sequence of distinct causal stages, each of which involves the voluntary action of some agent (or of a surrogate for human agency, such as a programmed machine), then it is permissible to adopt and act on this rule or procedure only if the actions it requires at every stage are justifiable to each person at that time.[25]

---

[24] Frick, "Contractualism and Social Risk." p. 202.

[25] Frick. p. 205. Although this helps the ex ante approach with some cases, it does not solve the general problem that the ex ante approach is inconsistent over time. See below. For a more detailed discussion, see Kacper Kowalzcyk, "Yet Another Argument Against Anti-Aggregation" (unpublished manuscript). pp. 28-33. I am indebted to Kowalzcyk's discussion of these issues.

So, if in a case like *Human Experiment*, we know that adopting a given plan would be disallowed by our own theory at a later stage, we should foresee that and thus not adopt the plan. In other words, we should take what, in dynamic contexts, is known as a *sophisticated* approach.[26] On this approach, the decision maker should act with foresight. We should look ahead at future choice nodes and take into account how we should decide in these choices. Then we use backward induction to reason from the very last choice node to our present choice node. In *Human Experiment*, this approach requires that we take into account the fact that at t_2 we would be acting impermissibly by going ahead with the experiment. We should not adopt a plan at t_1 that would require us acting impermissibly at a later stage. We should thus not run the lottery in the first place.

This to me seems the right way to reason about *Human Experiment*. But, of course, we could reason similarly about *Villain 2* and, in fact, about *Same Risk, Unknown Distribution*, even if there is no intervening agency at a later stage. I suggest that the fact that we are able to peek ahead and know that if we *were* to be given a second chance, we would want to overturn our choice, is enough to rule out (β) in *Villain 2* and in *Same Risk, Unknown Distribution* on the sophisticated approach.

I conclude that neither a resolute nor a sophisticated approach helps the proponent of the ex ante approach to get around the problems of violating Weak Stochastic Dominance.

We have now seen that the ex post approach can violate Weak Ex Ante Pareto in cases that are symmetric in the sense that there is an option that is in the best ex ante interest of all affected. We have also seen that

[26] See McClennen, *Rationality and Dynamic Choice*. Chapter 1.

ex ante contractualism, in turn, violates Weak Stochastic Dominance. This has become especially clear in Horton's *Villain 2*, which is a different risk case. So, we might ask whether there is a principled moral difference between different risk and same risk cases. It is this question I now turn to.

Same Risk vs Different Risk

Consider the following case.

> *Different Risk, Known Distribution*: The well-being of 10,000 X-people and 1,000,000 Y-people is at stake. We have to choose between α and β.
>
> (α) Only the Y-people are subjected to a dependent 99.99% risk of losing a leg.
>
> (β) Only the X-people are subjected to a dependent 10% risk of death.

This case is analogous to *Villain 2* in that it also is a different risk and a known distribution case. The numbers and probabilities involved, of course, are very different. Still, the same reasoning applies. The X-people are in a very different position from the Y-people. There is a clear conflict between what is in the Y-people's best interest ex ante and what is in the X-people's best interest ex ante. Each of the Y-people would want you to opt for (β), as this would be almost certain to leave them unharmed. The X-people on the other hand would want you to choose (α), as this would be sure to save their lives, whereas (β) would object them to a 10% risk of death each.

What is different from *Villain 2* is that in this case the ex ante and ex post approaches might pull in the same direction. If we assume that a complaint against being subjected to a 10% risk of death is stronger than a complaint against being subjected to a 99.99% risk of losing a leg, both would tell us to choose (α), as this would minimize both the strongest complaint ex ante and the strongest complaint ex post.[27] Next, contrast it with the following case.

> *Same Risk, Known Distribution*: The well-being of 1,010,000 people is at stake. We have to choose between α and β.
>
> (α) All are subjected to a dependent 99% risk of losing a leg.
>
> (β) All are subjected to a dependent 0.1% risk of death.

This, of course, is just the case we started out with, *Same Risk, Unknown Distribution*, except that the probabilities involved are dependent in this case. I will consider the question whether this plausibly makes a decisive difference below. For now, we can just note that all the same reasoning from above applies here as well. It is thus clear that the ex ante approach tells us to choose (β), whereas the ex post approach tells us to choose (α).

Suppose, however, that in *Different Risk, Known Distribution*, neither we nor the people themselves know who is an X-person and who is a Y-person. What would the situation look like in this case? Each would roughly face a 1% chance of being an X-person and a 99% chance of

---

[27] Of course, you might think that a 10% risk of death is not high enough to outweigh almost certainly losing one leg. We could, of course, run the same example with a higher risk. That would just make calculations more tedious later on. So, for the sake of argument, just assume that 10% is high enough, so that both approaches agree in this case.

being a Y-person.[28] This means that (α) would subject them to a 99% risk of losing a leg, while (β) would subject them to a 0.1% risk of death.[29] These, of course, are the exact same odds as in *Same Risk, Known Distribution*.

The initially clear-cut distinction between same risk and different risk cases starts to blur. By stipulating that people do not know whether they belong to the X or the Y group, we turn one category of case into another category of case. It is implausible to think that whether people know which group they belong to should matter to the moral decision maker. Moreover, even if it did matter, we can plausibly stipulate that in *Different Risk, Known Distribution* there was an earlier point in time where the people themselves did not know which group they are in. Making a decision *at that time*, we would be facing *Same Risk, Known Distribution*.

Thus, these two cases help us to see the consistency problems of the ex ante approach more clearly. Suppose at t_0, we are informed that we face *Different Risk, Known Distribution*. The ex ante approach tells us to choose (α). Suppose that at t_1 we find out that in fact, neither we nor the affected people know who is an X-person and who is a Y-person. This means that, in effect, we are facing *Same Risk, Known Distribution*. The ex ante approach now tells us to choose (β). Next, suppose that at t_2 we learn who the unlucky people are that are going to die. Invariably, we would now wish to be able to go back and overturn our choice yet again, opting for (α) instead. To me this seems implausibly

[28] The chance of being an X-person is $\frac{10{,}000}{1{,}010{,}000}$ or 0.99009…% and the chance of being a Y-person is $\frac{1{,}000{,}000}{1{,}010{,}000}$ or 99.009…%.

[29] The risk of losing a leg if α is chosen is $\frac{1{,}000{,}000}{1{,}010{,}000} \times 0.9999$, which is 99%. The risk of dying if β is chosen is $\frac{10{,}000}{1{,}010{,}000} \times 0.1$, which is 0.099…%.

unstable. In effect, whether a case is same risk or different risk only determines *when* a given conflict occurs. Prima facie at least, it is unclear why this feature should be of moral relevance.

Either through concealing the identities of the people involved or by assessing the decision from some other point in time, we can turn a different risk case into a same risk case. However, it is unclear why such information should be morally relevant. I conclude that it is not very plausible that these changes to the mere description of a case should change the *moral substance of a case*. This is a problem in particular for the ex ante approach, but it is also a general problem for theories that argue we should use fundamentally different approaches to same risk cases than we do to different risk cases. We will look at one such theory in section 2, below. But first, having investigated the distinction between same risk and different risk cases, I now move to the distinction between known distribution and unknown distribution cases.

Known Distribution vs Unknown Distribution

Whether a case is unknown distribution or known distribution determines whether the probability that a given conflict occurs can ever reach 100%. It is not clear that this particular feature is morally relevant. In particular, the fact that known distribution cases can be turned into unknown distribution cases very easily should make us doubt that there is any moral relevance to this distinction. For example, suppose that in

> *Russian Roulette*: A villain is playing Russian roulette on six people. He uses a gun with six chambers and loads it with one bullet. He points the gun at each of the six people in turn and pulls the trigger once.

This seems like a known distribution case, as we know that exactly one person is going to be killed. Suppose, however, that we know that with that specific gun, as with any gun I suppose, there is also a minuscule chance of it misfiring. This chance could be arbitrarily small. As long as it does not reach 0, we are dealing with an unknown distribution case. Should the introduction of such a tiny chance event fundamentally change the moral assessment of the case? I think not.[30]

More importantly, there is reason to believe that every risky case in reality is actually an unknown distribution case. It is very hard and most often irrational to be 100% certain about any future empirical fact. This implies that the very notion of a known distribution case might be an illusion in reality. But even if we do not go that far, it still seems implausible to hold that a step from a 99.9999…% risk to a 100% risk should make all the difference to how we should think about a case. I, therefore, do not think that the distinction between unknown distribution and known distribution cases *per se* can ground the claim that we should be using different approaches to either kind of case.[31]

Consider, however, the following case.

---

[30] Note, however, that the introduction of such a tiny chance event implies that the principle of Weak Stochastic Dominance might cease to apply, strictly speaking. However, we could slightly strengthen the principle to take care of this.

[31] For a similar thought, see also Otsuka, "Risking Life and Limb: How to Discount Harms by Their Improbability." pp. 87-89.

> *Known vs Unknown Distribution*: The lives of Ann and Bob are at stake. We have to decide between the following.
>
> (α) Subject them to a dependent 50% risk of death.
>
> (β) Subject them to an independent 50% risk of death.

Does it matter whether we choose (α) or (β) in this case? We might think not, as I have argued above that the distinction between unknown distribution and known distribution should not matter in itself. The ex ante approach agrees. Ann and Bob each have a complaint against (α) that is exactly as strong as their complaint against (β), namely a complaint against being subjected to a 50% risk of death. The ex post approach might disagree, however. If we choose (α), we know that exactly one of Ann and Bob is going to die. If we choose (β) on the other hand, things look different. There is a 25% chance that only Ann dies, a 25% chance that only Bob dies, a 25% chance that both die and a 25% chance that no one dies. This implies a 75% probability that at least one person would die if we choose (β). This is the quantity that matters for the ex post approach and we minimize it by choosing (β) rather than (α).[32] Remember that we want to minimize the *single* strongest individual complaint against our choice. For the ex post approach, this requires minimizing the chance of such a claim occurring.

This is why, on the ex post approach, we ought to choose (β) in the above case. This does not mean, however, that the ex post approach draws a principled distinction between unknown distribution and

---

[32] See section 3 below.

known distribution cases, much less that we should use different approaches to deal with either category.[33]

What is true however, is that in unknown distribution cases, there is always a chance that the harm, and therefore the corresponding complaint, does not occur. In cases with large numbers of people, as the cases above, this chance typically is small enough to be negligible. In a case like *Known vs Unknown Distribution*, however, the chance that no one would die and therefore no complaint would occur is quite significant. And since we are dealing with a non-aggregative theory, this chance is not outweighed by the chance that more than one will die, at least not obviously so.[34] We want to minimize the single strongest individual claim. In *Known vs Unknown Distribution*, this is done by choosing (β) rather than (α).

This concludes my discussion of the distinction between unknown distribution and known distribution cases. We have seen that it could plausibly sometimes make a difference to what we ought to do. But this is different from the claim that there is a fundamental moral difference between the two that would warrant using one approach for one and another approach for the other kind of case. This is what a hybrid account that makes a principled distinction between unknown

[33] Victor Tadros argues that out of concern for the separateness of persons, the imposition of harms resulting from known distribution cases might be harder to justify than harms resulting from unknown distribution case. See Victor Tadros, "Controlling Risk," in *Prevention and the Limits of the Criminal Law*, ed. Patrick Tomlin, Andrew Ashworth, and Lucia Zedner (Oxford University Press, 2013). I am not convinced by his argument, which is very speculative. However, I am indebted to Tadros' discussion of cases similar to *Known vs Unknown Distribution*.

[34] Of course, strict non-aggregation might be implausible for other reasons, but this is not my topic here. The purpose of this essay is to answer the question, how a non-aggregative theory is most plausibly extended to conditions of risk.

distribution and known distribution cases would have to claim, however. I do not think that such an account would be promising. I now move to discuss an initially more promising hybrid account that distinguishes between same risk and different risk cases.

II. A Hybrid Approach

We have seen that the ex ante approach violates Weak Stochastic Dominance and so is prone to be inconsistent over time, whereas the ex post approach violates Weak Ex Ante Pareto and so sometimes requires us to overrule the (implicit) consent of all involved. There are more objections against both approaches, but these are the most fundamental ones. Furthermore, we have seen that the distinctions between the different sorts of risky cases are not very clear cut and not obviously morally relevant.

This means that the challenge for any hybrid account that employs one approach for one kind of case and another approach for another kind of case, is twofold. First, the account needs to be able to satisfy both Weak Ex Ante Pareto and Weak Stochastic Dominance. For why bother going with a more complicated hybrid account when it is unable to solve the most pressing problem of either account? The second challenge is that it needs to show that there actually *is* a morally significant and clear-cut difference between the different kinds of cases that would license using different models for these different cases. Otherwise, there is no moral foundation for employing different approaches. In this section, I argue that it is very hard to meet either of these challenges and that a recent, initially promising, hybrid account fails them both.

To begin, note that, on the face of it, the idea of using the ex ante approach for what I called same risk cases and using the ex post approach for different risk cases is appealing. In same risk cases, like Frick's *Mass Vaccination (unknown victims)*, our intuitions lie with the ex ante approach, whereas in different risk cases, like Horton's *Villain 2*, our intuitions lie with the ex post approach. Therefore, the prospect of finding an account that lets us have our cake and eat it too, as it were, does sound attractive.

In a recent paper, Alec Walen attempts this task.[35] According to his account, we should sharply distinguish between "conflict-free" cases and "known-conflict" cases. We should take an ex ante approach to the former cases and an ex post approach to the latter cases. "Conflict-free" cases in Walen's terms are nothing else than what I above called same risk cases, while his "known-conflict" cases are what I called different risk cases. In Walen's words, in

> "known conflict" cases, either the agent or the patients, or both, can identify, at the time the agent must act, either by name or by definite description, which patients have which claims in a context in which the claims compete. […] In "conflict-free" cases, all of the patients, given what they know and could readily come to know, should want her to make one choice, and the agent, given what she knows and could readily come to know, has no reason to treat the case any differently.[36]

Walen takes the ex post approach to be the more basic approach. In some cases, however, he claims that we are justified in departing from this basic approach by employing the ex ante model. He writes that such a setup

[35] See Walen, "Risks and Weak Aggregation - Why Different Models of Risk Suit Different Types Cases."
[36] Walen. pp. 5-6.

> reflects the basic moral idea that we should care about individuals who will suffer harm to the extent that we believe that such individuals exist or will exist, *unless* there is sufficient reason to treat the case as a conflict-free one.[37]

This is Walen's hybrid account. Above, I claimed that such an account should meet two adequacy conditions. First, it should be able to satisfy both Weak Ex Ante Pareto and Weak Stochastic Dominance. Second, it should also be able to give us a justification for drawing a principled distinction between same risk and different risk cases. I now assess whether Walen's account meets these two tasks, starting with the first.

First, a hybrid approach that uses the ex ante approach for same risk cases *is* able to meet the requirement of satisfying Weak Ex Ante Pareto. This is because, by definition, this principle only applies to these cases. For Weak Ex Ante Pareto to apply, there has to be an alternative that is in the best ex ante interests of all affected people. This is the case in same risk cases, but not in different risk cases. In effect, this is the dividing line between these two categories.

Things are less clear regarding Weak Stochastic Dominance. I should note that at some points in the above, my analysis could have been taken to suggest that the ex ante approach violates Weak Stochastic Dominance only in different risk cases. That is not correct. It is only that such violations are particularly acute in such cases, like Horton's *Villain 2*. But in principle, the approach can violate Weak Stochastic Dominance in both kinds of cases. In fact, it does so in *Same Risk, Known Distribution*, which is akin to Frick's vaccination case, only that it is not a known distribution case. Here is this case again.

[37] Walen. p. 14.

Same Risk, Known Distribution: The well-being of 1,010,000 people is at stake. We have to choose between α and β.

(α) All are subjected to a dependent 99% risk of losing a leg.

(β) All are subjected to a dependent 0.1% risk of death.

In this case, as we have seen, the ex ante approach chooses (β). All involved individuals are in the same boat and we maximize their expected value by subjecting them to a 0.1% risk of death rather than letting them lose one leg with 99% probability. This is so, although we know that 1,010 people will in fact end up dead if we choose (β). There is an astronomically large number of possibilities of who exactly these people will be, but each of these possibilities, i.e. states of the world, is characterized by the following feature: 1,010 people are dead and the alternative would have been that no one is dead and instead 999,900 people lose one leg. Consider the following case.

No Risk: The well-being of 1,010,000 people is at stake. We have to choose between α and β.

(α) 999,900 identified people lose one leg.

(β) 1,010 identified people die.

In this case, every strictly contractualist approach would unequivocally opt for (α).[38] So, the ex ante approach will say that in every possible state of the world taken in isolation, (α) should be chosen over (β). Choosing (β) instead amounts to a violation of Weak

---

[38] Of course, by "strictly contractualist", I here mean "strictly non-aggregative." Again, whether such strict non-aggregation is plausible in general, is a different matter. More on this below in section 4.

Stochastic Dominance. This means that even a hybrid approach that confines the ex ante approach to same risk cases, still sometimes violates Weak Stochastic Dominance. So, one of the adequacy conditions I that claimed a hybrid approach needed to satisfy, is failed.

The second adequacy condition was that the account should be able to give us a justification for drawing a principled distinction between same risk and different risk cases. Walen suggests two ways to meet this condition. First, he refers to Frick's argument from the single person case.[39] I argued against this proposal at length above and will not iterate my objections here. The second suggestion is more promising. Walen claims that "people have the moral authority to waive their ex post claims."[40] This, Walen writes, in effect, is what we do when we, for example, give our consent to a medical procedure. We certify that we are aware of the risks, but still agree to the procedure because it is in our best interest ex ante. If things turn out badly, we are not justified in raising a complaint, because we waived it by agreeing to the procedure. In cases in which there is an action that in the rational self-interest of all involved, we can proceed *as if* all have waived these ex post claims.

This is an interesting proposal. However, we need to be clear that this does not salvage the ex ante approach, nor does it offer a principled defense of the hybrid approach. Rather, the suggestion amounts to the claim that although we should, in principle, take an ex post approach as a social planner, there are other things that we need to take into account, for example, the autonomy of people to take their own risks

---

[39] See Walen, "Risks and Weak Aggregation - Why Different Models of Risk Suit Different Types Cases." p. 17.
[40] Walen. p. 19

in life. So, instead of establishing a hybrid approach as the most plausible extension of contractualism and other non-aggregative theories to conditions of risk, this proposal holds that an ex post approach is the correct extension. However, by extending these theories to risk, we lay bare the fact that we need to bring in other resources to assist these theories in order to complete the moral picture.

I take this to be the main contribution of Walen's proposal. It does not establish the hybrid approach as an alternative to the ex ante and ex post approaches, but rather reemphasizes the point that both approaches are in some sense inadequate to account for all of morality.

I return to this suggestion in section 4 below. Let me now move on to developing the ex post approach.

III. The Ex Post Approach

As we have seen above, the ex post approach violates Weak Ex Ante Pareto. The ex ante approach, on the other hand, violates Weak Stochastic Dominance, which in turn threatens it to be inconsistent over time. We therefore face a clear choice. One of these prima facie compelling principles needs to give. This is a familiar choice from similar debates in the ethics of distribution, most notably the debate concerning ex ante and ex post variants of egalitarianism and prioritarianism.[41]

I find Weak Ex Ante Pareto less compelling than Weak Stochastic Dominance. The latter seems to me an indisputable principle of

[41] See e.g. chapter 7 of Adler, *Well-Being and Fair Distribution*.

rational choice for the reasons outlined above. The former, though intuitively appealing, seems less secure to me.[42] Given this clear choice, I side with Weak Stochastic Dominance and the ex post approach. So, what is the most plausible version of the ex post approach? This is the question I want to answer in this section.

The first pass at the ex post approach is what I call the naïve approach. To the best of my knowledge, this view is not defended by anyone in the literature but is sometimes used as a strawman to argue against *any* kind of ex post approach. The core of the naïve view is the claim that we should disregard uncertainty entirely. That is, we should reason about a case in which someone is subjected to a certain risk of harm exactly as they should reason about a case in which someone is subjected to that harm with certainty. This leads to the claim that, for example, Ann's complaint against being subjected to a 10% risk of being killed is just as strong as Bob's complaint against being killed with certainty. This claim is obviously implausible.

A slightly more plausible version of the naïve approach would hold that Ann's complaint is just as strong as Bob's complaint *if she actually dies*, i.e. the risk materializes, and non-existent if she does not die. This, we might call the *actualist ex post approach*. Under this proposal, we can only ever know if we made the right choice *after* the uncertainty is resolved. This is not very promising, as it would render the theory useless as a guide for moral action.

More promising is an approach by Sophia Reibetanz. She proposes an ex post approach that is very much in line with Scanlon's own original

[42] See also chapter 1 of this thesis for why I think that egalitarians should violate Ex Ante Pareto.

treatment of risk in *What we Owe to Each Other*.[43] According to Reibetanz,

> [a]s long as we know that [our choice] will affect someone in a certain way, we should assign that person a complaint that is based upon the full magnitude of the harm or benefit, even if we cannot identify the person in advance. It is only if we do not know whether acceptance of a principle will affect anyone in a certain way that we should allocate each individual a complaint based upon his expected harms and benefits under that principle.[44]

What this proposal highlights, is that, morally speaking, it seems to not matter *who* is burdened by a given action, as long as we know that *someone* is so burdened. This is an attractive idea that expresses a standard of impartiality that any plausible moral theory should fulfil. Her account seems to go in the right direction for the ex post approach. However, it is unclear why Reibetanz limits her proposal to cases in which we know (with certainty) that our choice will affect someone in a certain way.[45] In effect, this means that we should restrict the ex post approach to and known distribution cases and use the ex ante approach in unknown distribution cases. We have seen, however, that such a stark contrast between these two categories of cases is implausible.

However, we can extend Reibetanz's proposal. Only now, rather than resorting to an ex ante approach when we do not know with certainty whether a given complaint will arise, we take into account the

---

[43] See Scanlon, *What We Owe to Each Other*. e.g. p. 208 and Reibetanz, "Contractualism and Aggregation."

[44] Reibetanz, "Contractualism and Aggregation." p. 304.

[45] For similar criticisms of Reibetanz's proposal see Otsuka, "Risking Life and Limb: How to Discount Harms by Their Improbability."pp. 87-9 and Horton, "Aggregation, Complaints, and Risk." pp. 61-62.

probability that such a complaint will in fact arise. In this vein, Michael Otsuka writes that

> rather than following Reibetanz's advice, we should instead posit a complaint of premature death in this scenario, where this complaint is discounted in a manner that tracks the probability that *someone* would suffer this fate.[46]

This sounds right to me. Of course, in most of the cases we have been considering so far, the probability that at least one person will suffer a particular fate has always been (almost) 100%. To see the real value of Otsuka's proposal, we need to consider different cases. Take, for example, this scaled back version of *Same Risk, Unknown Distribution*.

> *Same Risk, Unknown Distribution 2*: The well-being of 1,000 people is at stake. We have to choose between α and β.
>
> (α) All are subjected to an independent 99% risk of losing a leg.
>
> (β) All are subjected to an independent 0.1% risk of death.

In the original version of this case, given the large number of people involved, it was statistically certain that at least one person would die if we opted for (β). In this version of the case, things look different. If we choose (β), there is a fairly high chance that no one dies. This probability is around 37%, which implies a corresponding 63% risk that at least one person dies.[47] According to Otsuka's proposal, we should thus posit a complaint against (β), which is a complaint against

---

[46] Otsuka, "Risking Life and Limb: How to Discount Harms by Their Improbability." p. 88, my emphasis.

[47] $\left(\frac{999}{1000}\right)^{1000}$ is the probability that no one dies and $1-\left(\frac{999}{1000}\right)^{1000}$ is the probability that at least one person dies.

death discounted to reflect this probability. Since it is statistically certain that at least one person will lose their leg if we choose (α), we then need to ask which is stronger, 100% of a complaint against losing a leg or 63% of a complaint against death. We can plausibly assume that in this case, the discounted complaint against (β) is still stronger than the undiscounted complaint against (α). However, if the risk decreases and the complaint is discounted even further, at some point our assessment would change and we would judge the complaint against (α) stronger than the heavily discounted complaint against (β). So, it is not, as the naïve approach would have it, that we disregard risk entirely.

There are, however, two problems with Otsuka's account. The first is what Otsuka calls the "constant marginal disvalue claim." This is the claim "that the moral significance of the difference between n versus n+1 individuals suffering a given harm is the same, for any whole number n, including the number zero."[48] Otsuka reaches this claim from reasoning about a case like *Known vs Unknown Distribution* that I introduced above.

> *Known vs Unknown Distribution*: The lives of Ann and Bob are at stake. We have to decide between the following.
>
> (α) Subject them to a dependent 50% risk of death.
>
> (β) Subject them to an independent 50% risk of death.

Above I claimed that, although it might be intuitive to be indifferent between both options, the ex post approach should favour (β), as this

[48] Otsuka, "Risking Life and Limb: How to Discount Harms by Their Improbability." p. 91.

minimizes the chance that *someone* will die. Otsuka, on the other hand, would claim that this smaller chance is balanced by the chance that both, Ann and Bob, die, which is 0 under (α), but 25% under (β). If we allow for such balancing, as Otsuka rightly notes, "[w]hat becomes relevant is simply the expected number who would suffer this harm under a given course of action."[49]

For Otsuka's account, this means that it cannot distinguish between (α) and (β) in my *Known vs Unknown Distribution*. This, we might think intuitively, is a desirable feature. However, it introduces an aggregative dimension, that most contractualists will be hesitant to accept, thereby violating the spirit of the ex post approach. The whole point of this approach is to extend a non-aggregative theory to conditions of risk.

How problematic this implication of Otsuka's account is, can be seen in a case like the following.[50]

> *Concentrated Risk:* The well-being of one X-person and one million Y-people is at stake. We have to choose between α and β.
>
> (α) The X-person loses one leg.
>
> (β) The Y-people are subjected to a perfectly correlated 0.0001% risk of death.

In this case, there are only two possible outcomes if we choose (β): Either all one million people die or no one dies. However, the *expected*

[49] Otsuka. p. 91.
[50] I was made aware such a case by Bastian Steuwer. See Bastian Steuwer, "Contractualism, Complaints, and Risk" (unpublished manuscript).

number of deaths and thus the expected number of complaints against death under (β) is 1. According to Otsuka's account, we would thus posit one undiscounted complaint against death. This complaint is stronger than a complaint against losing a leg, and so we should choose (α) over (β). However, what this overlooks is that the relevant question for the ex post approach is not, what the expected number of complaints is. Rather, it is the question how likely it is that *at least one person would have such a complaint*. We want to minimize the single strongest complaint ex post, irrespective of how many such claims there are. If we choose (β), this is 0.0001% of a complaint against death. If we choose (α) it is 100% of a complaint against losing a leg, and so we should choose (β) over (α). That Otsuka's account does not reach this judgement is the first problem of his account.[51]

The second problem stems from what I call *different harm cases*. In these cases, the harms that people face are heterogeneous. This means that the chance of a specific complaint arising might be very slim and therefore this complaint will be discounted almost entirely. However, the chance of a complaint of a certain magnitude arising might be very large, and this is what we want to take into account.[52]

[51] Of course, some may take this not to be flaw, but a virtue of the account. Intuitively, after all, we might think that we *should* choose (β) rather than (α) in *Concentrated Risk.* However, it is a problem of the account *insofar* as we want to employ it as an extension of strictly non-aggregative contractualism. This, of course, must not be what Otsuka himself has in mind.
[52] I thank Hilary Greaves for presenting this problem to me. I subsequently discovered that Joe Horton has raised it in print as well. See Horton, "Aggregation, Complaints, and Risk." pp. 65-6. I therefore do not spend much time on it.

We can deal with this problem by averaging the strongest complaints in each state of nature. This gives us the notion of the *expected strongest complaint* and we can now state

> *The Ex Post Approach*. It is permissible to choose prospect P over prospect Q if and only if the expected strongest complaint against Q is at least as strong as the expected strongest complaint against P.

I now turn to two objections against the ex post approach thus stated.

IV. Objections

In this last section, I reply to two important objections against the ex post approach. The first objection is that on the ex post approach, a series of individually permissible acts can turn out to be impermissible overall.

Sequential Consistency

According to Joe Horton, the ex post approach implausibly violates the following principle and should, therefore, be rejected.

> *Sequential Consistency*. If a moral view condemns act X, it should condemn any series of acts that is in all morally relevant respects equivalent to X.[53]

There are two ways to address Horton's objection. Either we can ask whether this principle is sound or we can ask whether the ex post

[53] Horton. p. 70. The label is mine, not Horton's.

approach actually violates the principle. Although both routes seem promising to me. I will here focus on the latter. Regarding the first route, let me just briefly mention the following. Although initially, this principle sounds appealing, it is well known that many theories violate something like this principle. For example, we know that it can be the case that if each individual in a group acts rationally, the group acts collectively irrational, and vice versa.[54]

Regarding the second route, we need to ask, even if the principle is sound, has Horton shown that the ex post approach is in violation of it? To answer this question, we need to ask whether his cases show that the ex post approach fails to condemn a series of acts that is *in all morally relevant respects* equivalent to an act that it does condemn. I do not think that Horton succeeds in showing that. To see this, we have to remind ourselves that the ex post approach is concerned with the magnitude of an alternative's expected greatest complaint. It thus minimizes a specific quantity. Only if this quantity is the same in two cases, yet the ex post approach treats these cases differently, Horton is on to something. This is not the case, however. Consider his

> *Villain 6*: A villain has kidnapped one person. We have to choose between α and β.
>
> (α) He will cut one hand off her.
>
> (β) He will run a lottery that gives her a one-in-a-billion chance of death.[55]

---

[54] For an overview of this and similar examples, see Part 1 of Parfit, *Reasons and Persons*.

[55] This and the following cases are taken from Horton, "Aggregation, Complaints, and Risk." p. 69. As before, I have slightly amended them for representational reasons.

By assumption, if we want to minimize the expected greatest complaint, we ought to choose (β). The quantity "complaint against losing a hand" is greater than the quantity "one billionth of a complaint against death." Consider next,

> *Villain 7*: A villain has kidnapped one billion children. We have to choose between α and β.
>
> (α) He will cut one hand off each child.
>
> (β) He will randomly select and kill one of the children.

Here, the ex post approach will choose (α). The expected greatest complaint against (β) is a full complaint against death, whereas the expected greatest complaint against (α) is a full complaint against losing a hand. Consider next,

> *Villain 8*: A villain has kidnapped one billion children. For each child in turn we have to choose between α and β. You know that, after you have chosen for all of the children, the villain will randomly select one of the one billion lottery tickets and then kill any child who has the corresponding ticket.
>
> (α) He will cut one hand off this child.
>
> (β) He will give this child a ticket for a lottery with one billion unique tickets.

It seems uncontroversial that we should treat every child in the same way. What we should do for one, we should do for all. Assuming this much, what would the ex post approach tell us to choose here? The expected greatest complaint against (α) is a full complaint against losing a hand. The expected greatest complaint against (β) is a full complaint against death. So, by the ex post approach, we ought to choose (α).

Thus, in *Villain 8*, on the ex post approach, we choose to prevent a full complaint against death over preventing a full complaint against losing a hand. Likewise, in *Villain 7* we choose to prevent a full complaint against death over a full complaint against losing a hand. And in *Villain 6* we choose to prevent a full complaint against losing a hand over preventing one billionth of a complaint against death. There clearly is no inconsistency here.

However, Horton's objection seems to be that in *Villain 8*, for each child, the ex post approach should choose (β), as this is just the same as choosing (β) over (α) in *Villain 6*. However, we cannot assume this, without assuming the falsity of the ex post approach. Therefore, Horton is begging the question against the ex post approach. It is not that in the one-person case we care for the probability that a particular individual dies, while in the multi-person case we care for the probability that someone dies. Rather, in both cases we only care for the probability that someone dies. It is only that in the one-person case these probabilities coincide. I therefore conclude that Horton's objection against the ex post approach does not succeed, even if Sequential Consistency is sound.

Everyday Life Cases and Ex Ante Pareto

This last objection to the ex post approach is arguably the most challenging. It holds that were we to employ the ex post approach, we would have to rule out many activities, that clearly seem permissible or even desirable. These are the considerations that drive us towards the ex ante approach. Frick's *Mass Vaccination (unknown victims)* is an example of such a case, but there are many others. Driving is often mentioned as an example of this sort. If billions of people take car trips for mere convenience, some of them are bound to seriously injure or kill some innocent person in an accident, even if the risk on each trip is minuscule. It therefore seems that the ex post approach would have to prohibit such activities. Still, intuitively, conscientious drivers do not act wrongly.[56]

There are (at least) four strategies available to proponents of the ex post approach to answer this objection. The first is to bite the bullet and to claim that although the ex post approach would prohibit these activities, it would rightly do so, as they actually *are* impermissible. The second is to claim that the ex post approach would actually not prohibit these sorts of activities. In other words, we could try to tell a debunking story. The third is to claim that although it is correct that the ex post approach would prohibit these activities *and* it would be wrong to do so, this does not undermine the ex post approach, but rather shows that we need to bring in other moral reasons to arrive at all things considered moral judgements. Relatedly, the fourth strategy is to let go the approach's insistence on disallowing interpersonal aggregation. I will take these four strategies up in turn.

I don't think that the first strategy is very promising. There might well be *some* cases that we intuitively think of as morally innocuous, but that are ruled out by the ex post approach and so we should revisit our

[56] See e.g. Kumar, "Risking and Wronging." pp. 27-28.

intuitions about these cases. Maybe this is the case for some activities that bring only small benefits now but are certain to cause great harm down the line. However, it is very implausible that this holds for all cases.

The second strategy is only slightly more promising. This reply holds that, in fact, the ex post approach does not prohibit the sorts of activities it is alleged to do. Take the example of driving again. First of all, one could argue that it seems false to claim that the (ex ante) benefits provided by driving are small compared to its (ex post) costs. One reason for this is that while contractualism forbids *inter*personal aggregation, it can acknowledge the moral significance of *intra*personal aggregation of benefits. And, as Johann Frick writes,

> [i]ntrapersonal aggregation may allow the contractualist to justify the performance of routine risky actions. While, on any given occasion, such a risky action may impose significant individual losses on some while providing only relatively minor benefits to others, over time almost everyone will come out ahead under a principle that permits risky actions of this type to be performed.[57]

Such reasoning lets the ex post approach judge routine risky activities, like driving, to be permissible. The individual accumulated benefits of allowing people to drive are great enough to warrant the risk it imposes on other people, even if we know that some of them will suffer down the line.[58]

I am doubtful, however, that this strategy is successful. First of all, it is not clear that even the aggregated benefits of convenience driving that

---

[57] Frick, "Contractualism and Social Risk." p. 179, fn. 8.

[58] For other reasons to doubt the "driving case", see Michael Ridge, "How to Avoid Being Driven to Consequentialism: A Comment on Norcross," *Philosophy & Public Affairs* 27, no. 1 (1998): 50–58.

an individual accumulates over a lifetime are great enough to outweigh serious injury or death to another person. Second of all, and more importantly, it seems clear that even if the debunking strategy would work for the ex post approach in this case, there are other cases, in which it wouldn't. Frick's *Mass Vaccination* and *Same Risk, Unknown Distribution* are cases in point. There are cases to which Weak Ex Ante Pareto applies and in which the ex post approach would recommend going against its recommendations. It is in these cases, in which the ex post approach seems to perform worst.

The third strategy for the ex post approach to respond to this objection is more principled and more promising. It is to argue that this shows that we need to take into account other reasons besides those given to us by the contractualist model. In other words, we scale back the ambitions of contractualism as a moral theory. This is Johann Frick's own proposed solution to deal with some of the problem cases for the ex ante approach.[59] According to this account, while contractualism provides us with one important class of moral reasons, there are other moral reasons we need to take into account as well. Sometimes these other reasons will outweigh the reasons given by the contractualist theory. In such cases, contractualism fails to provide an answer to the question of what is morally right, *all things considered*. This strategy can help both the ex ante approach and the ex post approach with some of their respective problem cases.

It helps the ex ante approach in a case like Horton's *Villain 2* by arguing that the strongest complaint ex ante is outweighed by considerations of overall value. It helps the ex post approach in a case like Frick's *Mass*

[59] See Frick, "Contractualism and Social Risk." pp. 219-223.

Vaccination, by claiming that the strongest complaint ex post is silenced by the fact that this claim has been waived in the manner that Walen suggested in the account I sketched in section 2 above.

I have previously argued that this ready resort to a pluralist account of interpersonal morality diminishes the attractiveness of contractualism as a standalone moral theory.[60] But maybe this is the right lesson to draw from trying to extend the theory to risk. However, it certainly does not speak against the ex post approach and *for* the ex ante approach. Rather, it is a lesson for contractualism as a whole.

I now turn to a fourth strategy for the ex post approach to deal with the objection of this section, which leads us in a similar direction. This is to drop its all-out ban on interpersonal aggregation. We can begin by reminding ourselves which cases the ex ante and ex post approaches "get right" intuitively.

Intuitively, the ex post approach yields the right answers in those cases in which it aligns with an aggregative approach, like Horton's *Villain 2*. Likewise, the ex ante approach gets those cases right in which it itself aligns with an aggregative approach, like in Frick's *Mass Vaccination* and *Same Risk, Unknown Distribution*. For example, to me it seems intuitively not wholly implausible that, uncertainty aside, we should let one thousand people die rather than let one million children live with only one leg for the rest of their life.[61] Recall the following case.

[60] See chapter 3 above.

[61] Of course, one million might not be enough. But as the number grows, I get the intuition that we should save legs rather than lives.

No Risk: The well-being of 1,010,000 people is at stake. We have to choose between α and β.

(α) 999,900 identified people lose one leg.

(β) 1,010 identified people die.

In this case, as I argued above, since there is no uncertainty, the ex ante and ex post approaches do not come apart. They both tell us to choose (α). Intuitively, however, it is not obviously clear that that is the right choice. A fully aggregative approach supports this intuition. Even a plausible approach of *limited* aggregation would likely tell us to choose (β).[62]

The counterintuitive implications of *both* the ex ante and the ex post extension of contractualism to conditions of risk are rooted in the contractualist ban on interpersonal aggregation. To get rid of these implications, both approaches need to accept at least a limited dose of aggregation. So, we should not think that we can settle the debate between proponents of the ex ante approach and proponents of the ex post approach just by looking at these implications. Rather, we should look at the internal structural features of the respective approach. The ex post approach does better in this regard than the ex ante approach.

Conclusion

In this essay I have argued for an ex post approach to extending contractualism (and other non-aggregative theories) to conditions of risk. I have done so in three ways. First, I have shown that we can

[62] For my own approach to Limited Aggregation, see Korbinian Rüger, "Aggregation with Constraints," *Utilitas*, forthcoming.

distinguish between risky cases along two dimensions: the same risk/different risk dimension and the and known distribution/unknown distribution dimension. I have argued that neither of these distinctions is significant enough to warrant using one approach to one kind of case and another approach to the other kind of case. I then rejected one such hybrid approach. Second, along the way, I have continued my earlier critique of the ex ante approach. This approach violates the plausible principle of Weak Stochastic Dominance and so is inconsistent over time. Under conditions of risk, contractualists ought to do as they know that they will want to have done. This requires taking an ex post approach. Third, I have outlined the most plausible version of the ex post approach and defended it against two important objections. I argued that one of these objections should be dealt with either by letting go of a strict ban on interpersonal aggregation, or by complimenting contractualism with other kinds of moral reasons. However, we should not take this to speak against the ex post approach.

5 | Impersonal Goodness, Personal Complaints

Many of our choices affect not only the well-being but also the identity and the number of people who will exist in the future. We can call such choices *variable population choices*. All existing moral theories have trouble assessing these kinds of choices. Traditionally, they deal with cases in which the identity and the number of people are held fixed across all alternatives, and only the well-being or other morally relevant properties of these people are assumed to be affected by our choice. It is unclear how existing moral theories can be extended beyond these *fixed population choices* without leading to implausible results in some cases.[1]

I here assess the merits of a particular pluralist account. I propose an approach to population ethics that aims to combine elements from a broadly consequentialist approach with elements from a broadly contractualist approach.[2] The former approach tells us to choose outcomes or distributions that maximize overall value.[3] The latter tells

[1] See e.g. Part 4 of Parfit, *Reasons and Persons*. Gustaf Arrhenius, *Population Ethics: The Challenge of Future Generations*, unpublished manuscript.
[2] In a recent article, Michael Otsuka pursues a similar project. See Michael Otsuka, "How It Makes a Moral Difference That One Is Worse off than One Could Have Been," *Politics, Philosophy & Economics* 17, no. 2 (May 2018): 192–215. I deal with Otsuka's account in depth below.
[3] I use the terms "outcome" and "distribution" interchangeably to describe the state of affairs an act results in. This includes the act itself. I also use the terms "general value" and "overall value" of an outcome interchangeably to refer to what is sometimes called the "all-things-considered value" of an

us to choose outcomes against which individual people cannot raise justified complaints. These two approaches often conflict in variable population cases. My goal is to resolve this conflict by arguing that both dimensions are important in these cases and neither approach can adequately deal with them on its own.

A contractualist complaints-based rationale, for example, offers a good explanation for the contention that it is more important to benefit existing people than it is to bring new people into existence, even if they would have good lives. However, in many cases, this approach will be silent, for example, when we have to choose between bringing one person and bringing another person into existence, without affecting anyone else. In these "non-identity cases," the complaints-based approach fails to offer any guidance. I argue that we therefore also need to take account of the overall or general value of possible outcomes. I shall argue, furthermore, that a plausible pluralist approach to population ethics that takes at least some of its reasons to stem from the general value of outcomes will in some cases be committed to the claim that we have an obligation to create people merely because their lives would be good. This requires us to give up on the popular *Neutrality Intuition*, according to which bringing people into existence that would have a life worth living is itself "morally neutral." I show, however, how such a pluralist approach to population ethics can account for a related notion, namely a weak version of the so-called *Procreation Asymmetry*.

I will proceed as follows. In the first section I will show how our intuitive judgements in certain cases lend support to a pluralist

---

outcome. Moreover, I do not distinguish between the value and the "goodness" of an outcome.

approach to population ethics that takes account of the individual complaints people can put forward against our choices and of the general value of the population that will exist as a result of our choice. In the second section, I distinguish between two different ways of making sense of this notion of general value*: person-affecting* and *impersonal*. I show what is at stake in this distinction by looking at the implications of either interpretation conditional on the response we give to another contentious question in population ethics. This is the question whether existence can be better or worse for a person than nonexistence. The affirmative answer to this question is a position we can call *Existence Comparativism*. In the third section, I will argue that the axiological construal of the Neutrality Intuition is false. We should let it go. On the sort of pluralist approach considered here, this also requires us to give up on the deontic construal of the intuition. In the fourth section, I introduce a recent proposal by Michael Otsuka along the lines sketched in the first section. I argue that it fails given my arguments of sections two and three. I then propose a different way of spelling out a pluralist approach to population ethics. In the fifth section, I extend the account by introducing cases with negative well-being. In some of these cases, my proposed approach yields counterintuitive results. I offer an amendment that does better in these cases and goes some way in dealing with another vexed issue in population ethics, the Procreation Asymmetry, by accounting for at least a weak version of it. In the sixth section, I respond to two objections, and in the seventh and last section, I conclude.

I. Pluralist Population Ethics

Suppose you have to choose between the following two options of bringing two people into existence. You can choose to bring about either outcome A, in which person *i* will lead a very good life and person *j* will lead a mediocre life, or outcome B, in which *j* will lead a very good life. In this second option, *i* will never exist. Instead of him, a different individual, *k*, will lead a life of the same quality as *j*'s in option A. This choice can be represented by the following table, where the numbers represent a person's overall quality of life or "lifetime well-being" in an outcome. I take this notion to be numerically measurable and interpersonally comparable. Moreover, for the purposes of this essay, I assume that the value of a single life is equal to the lifetime well-being of the individual living that life. I, therefore, exclude any factors other than well-being that could contribute to the value of a life. "Ω" represents the fact that a particular individual will never exist in a given outcome.

	i	***j***	***k***
A	10	5	Ω
B	Ω	10	5

Case 1

A and B are permutations of each other. Thus, under some moral theories, like utilitarianism, it does not matter, morally speaking, whether you choose A or B. This is implied by what Derek Parfit calls the *No-Difference View,* according to which the fact that *j* would exist in both outcomes makes no difference to the moral assessment of our choice.[4]

---

[4] See Parfit, *Reasons and Persons*. 366-9.

If we choose A, however, *j* can raise a complaint that we could have easily made her much better off. If we choose B, no one can raise such a complaint. *i* is not around and *k* could not have been better off. To me, it seems plausible that in this case we ought to choose B. We should deny the No-Difference View, at least in its deontic variant. There is a broadly contractualist justification for this choice. According to contractualism, for an action to be right, it must be justifiable to all. An action is justifiable to all just when it is licensed by a principle that cannot be reasonably rejected by any single individual.[5]

In Case 1, B is the only option that is justifiable to all, since A is not justifiable to *j*. We, therefore, have an attractive justification for choosing B, which to me is the intuitively correct choice. This approach, however, fails to explain our intuitive commitments in many other cases. Take, for example, the following case, which is a version of what Parfit has called the "Non-Identity Problem."[6]

	i	***j***	***k***
A	10	5	Ω
B	Ω	5	1

Case 2

In this case, it seems clear that we have an obligation to choose A.[7] This does not follow from the above reasoning, however. If we, against our best intuitive judgement, choose B, no one could raise a complaint that we could have made her better off, just as no one could raise such a

[5] Here and in the following I refer to Scanlon's Contractualism. See Scanlon, *What We Owe to Each Other*. especially pp. 189-248.
[6] See Chapter 16 of Parfit, *Reasons and Persons*.
[7] Following Parfit, most philosophers seem to agree with this assessment. For a recent dissenting voice, see David Boonin, *The Non-Identity Problem and the Ethics of Future People* (Oxford University Press, 2014).

complaint if we choose A. Accordingly, it seems we should, therefore, be indifferent and could permissibly choose B on the complaints-based approach.

Therefore, if choosing B is impermissible, the complaints-based rationale cannot be the whole story. Rather, it seems that some of our reasons in these cases stem from what we can call the *overall* or *general* value of the distributions we are choosing amongst, with A's value clearly exceeding B's value. Therefore, assessing which alternative has the greatest value is important for any plausible approach to population ethics.

A pluralist approach of the sort I am interested in would be able to deliver the verdict that we ought to choose B in Case 1 and that we ought to choose A in Case 2. It would say that we ought to choose B in Case 1 because the complaints people can put forward against our choices matter and that we ought to choose A in Case 2 because the general value realized by our choices also matters. In the next section, I deal with the question of how to interpret this notion of general value.

II. Which General Value?

If we accept that some of our reasons stem from the general value of a distribution, as I claimed we should, there is the further question, how the general value of a distribution should be interpreted. There are broadly two lines to take in answering this question. The first is an embrace of the notion of *impersonal* value of a distribution, which is then said to constitute the general value. According to this approach, we compare two alternative distributions not from the point of view of any individual in the distribution but from the "point of view of the universe." Though under most theories, impersonal value for the most part is *located* in the lives of individuals, this location in principle doesn't matter. This is the approach most commonly ascribed to G.E. Moore. As Donald Regan writes,

> [on the Moorean] approach, any agent's fundamental obligation is to bring about the existence of as much good as possible. The 'good' asserted here is an agent neutral good, not tied conceptually to any person's interests. But in fact the most important goods are certain kinds of events or experiences that occur within individual lives. Therefore, an important part of promoting the good turns out to be promoting certain events or experiences in individuals' lives.[8]

This approach holds that, in principle, we could ascribe value to a distribution above and beyond the value of its constituent lives. We could, for example, place impersonal value on the manner of distribution itself. This is done through choosing an appropriate aggregation function, that takes as its input the well-being levels of all lives in a distribution and combines them to produce the impersonal value of the whole distribution. The simplest aggregation function is the total utilitarian one that simply sums up all individual well-being levels. However, we could also have a prioritarian, egalitarian or some other more complicated function that places value on the pattern of distribution. According to such functions, the impersonal value of a distribution can vary, even if the sum of the individual well-being levels does not.

[8] Donald H Regan, "Why Am I My Brother's Keeper?," in *Reason and Value – Themes from the Moral Philosophy of Joseph Raz*, ed. R Jay Wallace et al. (Oxford: Oxford University Press, 2004), 202–30. p. 204.

The second approach to interpreting the notion of general value is a denial of what I have called impersonal value. According to this approach, the only kind of value that exists is *personal* value. The general value of a distribution must be reducible to *person-affecting* considerations. However, there is no uniform way of making this notion of "person-affectingness" precise. Different philosophers employ it in different ways. I here briefly sketch three ways of how to understand this notion, increasing in the strength of the restrictions they put on our understanding of general value.[9]

The first is the idea that general value supervenes on personal value. That is, there can be no change in general value without an underlying change in personal value. For the purposes of this essay, I am assuming that well-being is the only thing that determines the personal goodness or value of a life. This gives us the following

> *Weak Person-Affecting Restriction.* General value supervenes on individual well-being.

This is quite a weak understanding of the restriction. It amounts to a standard understanding of (evaluative) welfarism according to which the well-being of individuals is ultimately the only thing of value. A stronger understanding of the restriction is the following.

[9] Note that these are by no means the only ways to understand "person-affectingness." For a slightly different way to carve up the landscape of person-affecting ideas, see Gustaf Arrhenius, "Can the Person Affecting Restriction Solve the Problems in Population Ethics?," in *Harming Future Persons*, ed. David Wasserman and Melinda A. Roberts (Springer, 2009), 289–314.

Strong Person-Affecting Restriction. If a distribution X is better than another distribution Y, then X is better than Y for at least one individual.[10]

This is a stronger understanding of the restriction, and arguably the most common one in the population ethics literature. It corresponds to what Larry Temkin has called "the Slogan", which states that "[o]ne situation cannot be worse (or better) than another if there is no one for whom it is worse (or better)." [11] Temkin claims that the Slogan underlies much of contemporary moral thinking.

To see the difference between this and the first version of the restriction, consider the case of an extreme form of egalitarianism, according to which equality in well-being is the only thing that matters for a distribution's value. On such a theory, we could make an unequal distribution better by "levelling down" or even making everyone worse off. Such a theory is excluded by the Slogan, but not by the first interpretation of the person-affecting restriction. After all, even on such a theory, it is still the case that general value supervenes on the well-being of individual people. Without a change in individual well-being, there is no change in the general value of a distribution. Nonetheless, such extreme forms of egalitarianism are usually not understood as person-affecting theories, and rightly so, I think. Normally, when we say that morality is person-affecting, we mean more by that. We mean something like the Slogan.

[10] This corresponds to what Parfit calls the "narrow telic principle." See. Derek Parfit, "Future People, the Non-Identity Problem, and Person-Affecting Principles," *Philosophy & Public Affairs* 45, no. 2 (March 2017): 118–57. p. 118. For now, I set aside Parfit's "deontic" and "wide" principles.

[11] Larry S. Temkin, *Inequality* (Oxford University Press, 1993). p. 248.

For some theorists, however, even this restriction does not go far enough. For example, Ralf Bader writes that it

> is a rather weak necessary condition on betterness orderings that can be satisfied by impersonal theories that merely assign ethical relevance but not ethical significance to personal good. The positive commitment of a person-affecting view […] is the commitment to the reducibility of general good and the rejection of impersonal good.[12]

For this even stronger interpretation of the person-affecting restriction, anything that we can say about the value of a distribution has to be *reducible* to facts about the personal value for individuals. This interpretation is stronger than the other two interpretations we have considered so far. It is, however, not easy to make precise.

One way of making it more precise, is by considering a theory that would be ruled out by it, but not by the two weaker interpretations of the person-affecting restriction. One example is the theory of prioritarianism. According to prioritarianism, an individual's well-being matters more/contributes more to the general value of an outcome, the worse off a person is.[13] As I understand prioritarianism here, the best outcome is the one maximizing the sum of priority-weighted moral value of well-being, where this notion in turn is reached by applying a strictly increasing concave function to well-being. Prioritarianism, thus understood, is ruled out by the following person-affecting restriction.

---

[12] Ralf M. Bader, *Person-Affecting Population Ethics*, unpublished manuscript. p. 43, n. 1.
[13] For the seminal statement of "the priority view", as he calls it, see Parfit, "Equality and Priority."

> *Very Strong Person-Affecting Restriction.* If a distribution X is better than another distribution Y, then X is better than Y for at least one individual *to a greater degree* than Y is better than X for any individual.[14]

To see how this restriction rules out prioritarianism, consider the following example.

	i	j
A	8	2
B	10	1

Prioritarian Case

In this case, whatever we choose, our choice will not be ruled out by the Slogan. This is because A is better than B for *j*, and B is better than A for *i*. So, each outcome is better than the other for someone. However, according to the Very Strong Person-Affecting Restriction, A cannot be better than B in this case. This is because, A is better than B for *j* to the degree of one unit, whereas B is better than A for *i* to the degree of two units. So, it cannot be the case that A is better than B.

Prioritarianism, however, tells us otherwise. Given a suitable conversion function from well-being (w) to prioritarian weighted moral value of well-being (V), e.g. $V = \sqrt{w}$, prioritarianism will judge A to be better than B. So, the Very Strong Person-Affecting Restriction rules out prioritarianism and considers it an impersonal theory, even in fixed-population cases. Utilitarianism, on the other hand, necessarily fulfils this restriction in fixed-population cases.

[14] For a similar restriction, see Ingmar Persson, "Equality, Priority and Person-Affecting Value," *Ethical Theory and Moral Practice* 4, no. 1 (2001): 23–39.

We have now looked at three specifications of the person-affecting restriction, from weak to very strong, to demarcate the person-affecting construal of general value from its impersonal counterpart. The weak restriction only rules out theories that aren't fundamentally based on individual well-being. The strong restriction in addition rules out extreme egalitarianism. The very strong restriction also rules out prioriarianism.

So, which is the "correct" Person-Affecting Restriction? I will not attempt to definitively settle this question here, and doubt that it can be settled. For the purposes of this essay, I think it is best to go with the strongest possible interpretation of the restriction. This helps us to see the distinction between person-affecting and impersonal value the most clearly. Not much of what I say in the following hinges on this assumption. However, it lets us clearly classify distribution-sensitive axiologies, like prioritarianism and egalitarianism, as impersonal axiologies, even in fixed-population cases. On the other hand, whether utilitarianism is a person-affecting or an impersonal axiology can only be determined once we move beyond fixed-population ethics. Consider, for example, the following case.

	i	j
A	9	Ω
B	9	10

Case 3

Impersonally speaking, B is clearly better than A. By choosing B, we realize a much greater amount of well-being than by choosing A. B also has higher average well-being.[15] In person-affecting terms, things are

[15] Of course, the inequality in B is higher than in A. This means that on some impersonal theories, like extremist egalitarianism, A would be better than B.

much less clear. This is because it is unclear whether B has any person-affecting considerations in its favour. This depends on whether it is better than A for *j.* If that were the case, then we could say that B is much better for *j* than A, and since *i* fares the same in both A and B, B is better than A overall. However, whether we can plausibly claim that B is better for *j* than A, i.e. whether life at well-being level 10 is better for *j* than nonexistence, is one of the most contentious questions in population ethics. I shall now move to discuss this question.

Comparativism

If we accept what we can call *Existence Comparativism* (or *Comparativism* for short), then we do think that it makes sense to claim that life at well-being level 10 is better for *j* than nonexistence.

> *Comparativism.* If person *p* exists at a positive (negative) well-being level in distribution X and does not exist in distribution Y, then X is better (worse) for *p* than Y.

If we accept Comparativism, we could then reach the judgement that in Case 3, B is better than A on purely person-affecting grounds. This would also help us to reach the intuitively correct verdict in Case 2. Recall that case:

	i	***j***	***k***
A	10	5	Ω
B	Ω	5	1

Case 2

We could avoid this difficulty, by setting *j*'s well-being in B to 9. Then there would be no inequality.

If Comparativism is correct, we could plausibly say that A is better than B for *i*, *to a greater extent* than B is better than A for *k*, without having to leave the person-affecting framework. Comparativism, however, is very controversial. Jan Narveson puts the controversy as follows.

> If you ask, " whose happiness has been increased as a result of his being born? ", the answer is that nobody's has. […] Remember that the question we must ask about him is not whether he is happy, but whether he is happier as a result of being born. And if put this way, we see that again we have a piece of nonsense on our hands if we suppose that the answer is either " yes " or " no". For if it is, then with whom, or with what, are we comparing his new state of bliss? Is the child, perhaps, happier than he used to be before he was born? Or happier, perhaps, than his alter ego? Obviously, there can be no sensible answer here.[16]

John Broome agrees and thinks that Comparativism cannot be true.

> [I]t cannot ever be *true* that it is better for a person that she lives than that she should never have lived at all. If it were better for a person that she lives than that she should never have lived at all, then if she had never lived at all, that would have been worse for her than if she had lived. But if she had never lived at all, there would have been no her for it to be worse for, so it could not have been worse for her.[17]

Broome and Narveson both certainly make a strong case for Noncomparativsm. If they are correct, however, and Comparativism cannot be true, then a person-affecting account will not be able to reach the judgement that A is better than B in Case 2, at least not straightforwardly so. Broome's argument has been contested,

---

[16] Jan Narveson, "Utilitarianism and New Generations," *Mind* 76 (1967): 62–72. p. 67.

[17] John Broome, *Ethics out of Economics* (Cambridge: Cambridge University Press, 1999). 168

however.[18] One line of criticism concerns the fact that Broome's argument seems to presuppose that to claim that existence is better for person *p* than nonexistence is to express a *triadic* betterness-relation between a state in which *p* is alive (call this S1), a state in which *p* isn't alive (call this S2) and *p* herself.

It is not clear, however, that the betterness relation between existence and non-existence actually is a triadic relation. One could hold instead, rightly so I believe, that the betterness relation we are interested in is between *p's* life in S1 and *p's* life in S2. This relation between the two possible lives, however, is dyadic rather than triadic. If the set of all possible lives that could be lived by *p* includes a "null life," corresponding to never being born, then there seems to be nothing incoherent about Comparativism.[19] If on the other hand, however, the set of possible lives does not include such a null life, then we are back to the incoherence argument, only now in terms of a dyadic relation where one relatum is missing if the person does not exist in both outcomes under consideration.[20]

This second option seems more plausible to me. The notion of a "null life" as a member of the set of all possible lives seems to me metaphysically dubious. In effect, this would amount to ascribing a well-being level of zero to a person in a state in which this person does

[18] See e.g. Nils Holtug, "On the Value of Coming into Existence," *The Journal of Ethics* 5, no. 4 (2001): 361–384. Gustaf Arrhenius and Wlodek Rabinowicz, "The Value of Existence," *The Oxford Handbook of Value Theory*, 2015, 424–443. Marc Fleurbaey and Alex Voorhoeve, "On the Social and Personal Value of Existence," in *Weighing and Reasoning: Themes from the Philosophy of John Broome*, ed. Iwao Hirose and Andrew Reisner (Oxford University Press, 2015), 95–109.
[19] I thank Hilary Greaves for raising this point to me.
[20] For a defense of this idea, see chapter 1 of Bader, *Person-Affecting Population Ethics*. I am indebted to Bader's discussion in what follows.

not exist. As Arrhenius and Rabinowicz rightly note, "it is quite nonsensical to ascribe any well-being level at all to a person in a state in which she does not exist. Well-being presupposes being."[21]

To avoid these metaphysical pitfalls, Arrhenius and Rabinowicz propose a different approach. They propose an approach that lies somewhat in between Comparativism and Noncomparativism.[22] According to this *Restricted Comparativism*, Broome is wrong in claiming that from the claim that it is better for *i* to lead a good life than never existing at all, it follows that never existing would have been worse for *p*. Instead, they argue (also presupposing a triadic betterness relation) since people only exist when they are alive, one relatum is missing when *p* does not exist.

> Consequently, even if it is better for p to exist than not to exist, assuming she has a life worth living, it doesn't follow that it would have been worse for *p* if she did not exist, since one of the relata, *p*, would then have been absent. What does follow is only that nonexistence is worse for her than existence […] but not that it would have been worse if she didn't exist.[23]

According to this account, in the above Case 3, we can say that *conditional on choosing B*, B is better for *j* than A, but choosing A would not have been worse for *j*. According to Arrhenius and Rabinowicz, we can and should reject the following principle, which is implicit to Broome's argument:

---

[21] Arrhenius and Rabinowicz, "The Value of Existence." p. 429.
[22] See Arrhenius and Rabinowicz.
[23] Arrhenius and Rabinowicz. 428.

> *Subjunctive Connection 1 (SC1)*. An outcome A is better (worse, equally as good) for p than (as) another outcome B only if outcome B would be worse (better, equally as good) for p than (as) A if B came about.[24]

Instead, we could accept the following principle:

> *Subjunctive Connection 2 (SC2).*
>
> (i) If a person p exists in both outcomes A and B, then A is better (worse, equally as good) for p than (as) B only if B would be worse (better, equally as good) for p than (as) A, if B obtained.
>
> (ii) If a person p exists in A but not in B, then A can be better (worse, equally as good) for p than (as) B although B would not be worse (better, equally as good) for p than (as) A, if B obtained.[25]

I am very much drawn to the stronger SC1. I *do* think that from the claim that outcome A is better for *p* than outcome B, it follows that B would be worse than A for *p*, if B obtained. However, I do not have a knock-down argument for that.[26] Maybe Arrhenius and Rabinowicz are right and SC1 only seems attractive because it is "more in line with our common way of thinking," since we are "accustomed to compare outcomes in both of which the affected person exists."[27] So, for the sake of argument, let us consider Restricted Comparativism a live option for now.

Concerning the Existential Question, we now have two options left on the table: Noncomparativism and Restricted Comparativism. What

---

[24] Arrhenius and Rabinowicz. p. 428.
[25] Arrhenius and Rabinowicz. p. 429.
[26] Though I do believe that such an argument is not too far away. See e.g. Krister Bykvist, "The Benefits of Coming into Existence," *Philosophical Studies* 135, no. 3 (September 8, 2007): 335–62. pp. 50-3 and chapter 1 of Bader, *Person-Affecting Population Ethics*.
[27] Arrhenius and Rabinowicz, "The Value of Existence." p. 429.

would the implications of these two positions be, when coupled with a strictly person-affecting axiology? This shall be the topic of the next two subsections.

Person-Affecting Axiology with Noncomparativism

We could deny both impersonal value and Comparativism. This would imply that anything that is better or worse about a distribution must be better or worse for someone and that it is not better for someone to have a good life than never existing. This combination of an insistence on person-affecting value and a denial of Comparativism, however, seems to lead to clearly unacceptable implications. Recall Case 2.

	i	*j*	*k*
A	10	5	Ω
B	Ω	5	1

Case 2

It seems that strict Noncomparativism paired with a person-affecting axiology could not reach the judgement that A is better than B, for it is not better than B from any personal point of view.

For me, this is enough to reject the pairing of person-affecting general value and Noncomparativism.[28] Ralf Bader has proposed a theory, however, that, according to Bader, allows us to take care of this case even when sticking to our insistence on person-affecting value and a denial of comparativism. The key to this proposal is that the personal

[28] As I note above, however, some are prepared to reject the "non-identity intuition" in cases such as this. See Boonin, *The Non-Identity Problem and the Ethics of Future People*.

points of view, to which we are confined within the person-affecting framework, need not be occupied by the same individuals in each distribution. This stems from a plausible commitment to *Impartiality* or *Permutation Invariance*. This condition, which is often also called *Anonymity*, holds that switching the identities of people in a distribution does not change the value of the distribution if everything else is held fixed.[29]

> *Anonymity (axiological).* If two distributions X and Y are permutations of each other, then X and Y are equally good.

Anonymity is an intuitively attractive principle. For Bader's theory, accepting Anonymity means that we can compare partially distinct and entirely distinct populations without giving in to an impersonal assessment from the "point of view of the universe" or the "view from nowhere." Rather,

> [a] commitment to impartiality implies that a situation can be better than another even though there is no one for whom this is better. Since an impartial ordering holds that the identities of people are not relevant for betterness claims, it is enough that there is someone in the one situation who is better off than someone else in the other situation.[30]

It is unclear, however, whether this proposal classifies as a person-affecting theory in the above sense. Strictly speaking, Bader's theory violates what I called the strong and very strong person-affecting principles above. These principles are sensitive to the identities of the

[29] See e.g. Broome, *Weighing Lives*. pp. 135-6.
[30] Bader, *Person-Affecting Population Ethics*. p. 71.

affected individuals, in a way that Bader's theory is not, because of its insistence on Anonymity.[31] Bader himself writes that the theory

> nevertheless classifies as a person affecting theory insofar as it evaluates distributions exclusively in terms of what is good for persons, namely in terms of personal good.[32]

I am unsure about this classification. I suspect that most people that are attracted to person-affecting morality would reject it. However, if we grant it for the moment, Bader's proposal would mean that we can, at least, compare all distributions that are exactly equally large without giving up on the person-affecting framework. The theory, however, has problematic implications in different number cases. Because of its rejection of comparativism, it deems all distributions of different size incomparable in terms of general value. As Bader writes, it

> imposes an equinumerosity constraint on the general betterness relation. Distributions can be compared if there are bijections that map the members of the one distribution onto their corresponding images in the other distribution.[33]

This restriction to same number cases is very problematic. Take the following case.

	i	***j***
A	1	Ω
B	10	10

Case 4

In this case, Bader's theory would deem A and B incomparable. This by itself is a problematic feature, as B seems clearly better than A,

[31] In this sense, Bader's theory is similar to Parfit's *wide* person-affecting principles. See Parfit, *Reasons and Persons*. pp. 396-401.
[32] Bader, *Person-Affecting Population Ethics*. p. 71.
[33] Bader. p. 75.

intuitively. B is much better than A for *i.* So, it seems that even on purely person-affecting grounds we would want to say that B is better than A. The theory seems even more problematic when we consider the following three-option case.

	i	***j***
A	1	Ω
B	10	10
C	Ω	10

Case 5

According to Bader's theory, C is better than A, but A and B are incomparable. This is very implausible insofar as C is transformed into B by giving *i* 9 more units compared to what he would have had in A. So, if C is better than A, the only intuitively plausible inference would be to hold B better still. Bader's theory, however, does not have the resources available to make this judgement. I take this to be a very substantive problem for the theory. The theory is silent in many cases, where we intuitively think it should be very easy to make comparative value judgements.

Bader's theory is an improvement on a more straightforward noncomparativist and person-affecting theory, as it can at least give us same-number comparisons. However, widespread incompleteness remains. The "equinumerousity restriction" means that when we compare two populations of billions and billions each, if one population has so much as one person more in it than the other, the comparison fails. As Hilary Greaves writes, "there may well be some incompleteness in the betterness relation, perhaps especially in variable-population cases, but the incompleteness cannot plausibly be as radical as this."[34] This should lead us to reject person-affecting axiologies that insist on strict Noncomparativsm. What about such axiologies that allow for the above sketched restricted form of Comparativism?

Person-Affecting Axiology with Restricted Comparativism

If they are not wedded to strict Noncomparativism, person-affecting theorists could combine their rejection of impersonal value with the acceptance of Restricted Comparativism. While such a view does not (straightforwardly) yield the widespread incomparability in different number cases as above, it still struggles to make sense of value comparisons that should be easy to make. Recall Case 2

	i	***j***	***k***
A	10	5	Ω
B	Ω	5	1

Case 2

The view under consideration would be able to judge A better than B, only if outcome A is actual. For, according to this view, only if *i* exists, it makes sense for us to say that A is better than B *for i*. This feature gives rise to a problematic implication. Reconsider Case 3.

[34] Hilary Greaves, "Population Axiology," *Philosophy Compass* 12, no. 11 (November 2017): 1–15. p. 12. Note that Bader introduces the notion of "meta-betterness" to partly deal with the widespread incompleteness his theory gives rise to. See Bader, *Person-Affecting Population Ethics*. pp. 94-99. I am unsure how promising this proposal is and do not have the space to discuss it here in detail. However, it seems clear to me that in Case 5, B is better than A in terms of overall or general value, and not just "meta better."

	i	*j*
A	9	Ω
B	9	10

Case 3

In this case, according to the position under consideration, if we bring about B, then B is better for *j* than A, and thus better overall. If we bring about A, then A is not worse than B for *j*, and thus it is not the case that B is better than A overall. This violates the following principle.

> *Axiological Invariance*. The value of an outcome does not depend on whether it is brought about.[35]

This principle is intuitively plausible. Moreover, a violation of it is especially troublesome when we assume, as I do in this essay, that our axiology provides normative reasons. If our axiology violates invariance in the above sense, then it might provide us with conflicting reasons for action. What we ought to do might depend on what we actually end up doing. As Bykvist writes,

> When you use a theory as a guide to action, you use the theory in your deliberations about what to do. On the basis of this deliberation you then make up your mind and decide what to do. But if an action's rightness depends on whether it is performed, then in order to decide whether an action is right you first have to know whether or not you are going to perform it. But there is no point in deliberating about whether to perform an action if either you believe

[35] This is an axiological variant of a more common principle, *Normative Invariance*. For this principle see. e.g. Erik Carlson, *Consequentialism Reconsidered* (Kluwer Academic Publishers, 1995). p. 100. Note that this principle refers to the general value of an outcome and not just the value for a particular person. For this, more restricted, use see e.g. Krister Bykvist, "Being and Wellbeing," in *Weighing and Reasoning: Themes from the Philosophy of John Broome*, ed. Iwao Hirose and Andrew Evan Reisner, First edition (Oxford ; New York, NY: Oxford University Press, 2015), 87–94. p. 88.

> that you will perform it, or you believe that you will not perform it.[36]

In short, by violating Axiological Invariance, a theory that recognizes axiological reasons could cease to be action guiding. Thus, I don't think that a plausible moral theory could include an axiology of this structure. For me, this is enough to reject Restricted Comparativism. However, I will not set it aside completely just yet. It will resurface in my discussion of the Neutrality Intuition below.

This concludes my discussion of person-affecting axiologies for now. We have seen that an insistence on person-affecting value coupled with strict Noncomparativism leads to implausibly widespread incompleteness. On the other hand, coupling person-affecting value with Restricted Comparativism gives rise to a problematic normative structure. All of this casts doubt on the promise of insisting on person-affecting value. Before we move on to impersonal value, however, let me pause to introduce a different approach that does not neatly fit into one of the categories mapped out here. This approach invokes the notion of a "noncomparative benefit" and was recently proposed by Jeff McMahan.

Noncomparative Benefits

As I claim above, Case 1 shows us that the No-Difference View as a judgement about permissible choice is false. Here it is again.

---

[36] Bykvist, "The Benefits of Coming into Existence."pp. 350-351.

	i	*j*	*k*
A	10	5	Ω
B	Ω	10	5

Case 1

We should not be indifferent between A and B. Rather, we ought to choose B over A. As in Case 1, we can explain this by invoking the notion of a complaint that *j* can raise if we fail to make him as well off as he could have been. Choosing A is not justifiable to *j*. It is important to note that this leaves open the possibility that the No-Difference View is correct as a purely axiological claim. In fact, I think it is. I believe that the outcomes generated by A and B *are* equally good. This by itself does not imply that we should be indifferent between them. To be sure, the fact that two outcomes are equally good gives us no reason to prefer one to the other and if there are no other reasons at play, then we should in fact be indifferent. In this case, however, there seems to be another reason at play. The fact that choosing A would give rise to a complaint and choosing B would not, gives us a reason to prefer B. This is the whole point of a pluralist proposal to population ethics that reconciles axiological reasons with other normative reasons.

There is, however, one recent proposal that seeks to deny the No-Difference View without invoking the notion of complaints or other deontic considerations. This is a view proposed by Jeff McMahan, which Parfit calls *The Two-Tier View*.[37] According to this approach, we can distinguish between "ordinary" and "existential" benefits. Ordinary benefits are benefits that go to people who exist

[37] See Jeff McMahan, "Causing People to Exist and Saving People's Lives," *The Journal of Ethics* 17, no. 1–2 (June 2013): 5–35. and Parfit, "Future People, the Non-Identity Problem, and Person-Affecting Principles."

independently from the conferral of that benefit. Existential benefits, on the other hand, are benefits to which the beneficiary owes her existence. Most ordinary benefits are comparative, like the benefit of 5 to *j* if we choose B rather than A in Case 1. By doing so, we clearly make *j* better off. Most existential benefits, on the other hand, are noncomparative, like the benefit of 10 to *i* if we choose A. This benefit is noncomparative because it is not the case that by choosing A we make *i better off* than she would have been had we chosen B. For that would imply that had we chosen B, *i* would have been worse off. But this implication is false, as we have seen, if we don't want to accept the metaphysically questionable straightforward Comparativism. Had we chosen B, *i* would not have been at all.

However, as Derek Parfit writes, "[w]e may admit that for this reason, causing someone to exist cannot be *better* for this person. But it may be *good* for this person."[38] Following Parfit, McMahan holds, we still benefit *i* by choosing A, because A is *good for i* in the sense that he has a life in which, by assumption, the intrinsically good elements outweigh the intrinsically bad elements.[39] I am not as confident as McMahan that the use of the term "benefit" really is appropriate here, but, conceptually, it seems straightforward to make sense of this idea, which McMahan calls a *noncomparative existential benefit.*

McMahan goes on to suggest that comparative benefits might matter more than noncomparative benefits.[40] On this account, if we can choose between conferring a comparative benefit and conferring a noncomparative benefit of the same magnitude, we should choose the

[38] Parfit, *Reasons and Persons*. p. 489.
[39] See McMahan, "Causing People to Exist and Saving People's Lives." p. 6.
[40] See McMahan. pp. 14-15.

former. Like a complaints-based approach, this can explain why we should choose B in Case 1. The approach thereby also denies the No-Difference View. What distinguishes the two approaches is the way they arrive at this judgement. The main difference, as pointed out, is that a complaints-based approach only denies the deontic No-Difference View. The Two-Tier View denies both the deontic and the axiological version of the No-Difference View. As McMahan writes, on this proposal,

> an existential benefit contributes less to making the outcome better than an equivalent ordinary benefit and […] there is a correspondingly weaker reason to bestow the existential benefit than there is to bestow the equivalent ordinary benefit.[41]

Since I believe that the axiological No-Difference View is correct, I take this to be an advantage that the complaints-based approach has over the Two Tier View. The main reason why I think that the No-Difference View should not be rejected on axiological grounds, but purely on deontic grounds, is that to reject it on axiological grounds, we would have to accept views about the structure of the good that are deeply problematic. We would have to accept something along the lines of what Larry Temkin calls the *Essentially Comparative View of Outcome Goodness*.[42] According to this view, how good an outcome is, may depend on the other outcomes we are comparing it to.[43]

---

[41] McMahan. p. 15. See also Parfit, "Future People, the Non-Identity Problem, and Person-Affecting Principles."

[42] See Temkin, *Rethinking the Good : Moral Ideals and the Nature of Practical Reasoning*. 371.

[43] See also Johann Frick, "Context-Dependent Betterness and the Mere Addition Paradox" (unpublished manuscript). Frick argues that though Temkin is wrong in claiming that the value of outcomes is "essentially comparative," such value can be "context dependent." Though more plausible than the essentially comparative view, Frick's proposal gives rise to similar structural problems.

To deny that in Case 1, A and B are equally good, which amounts to the axiological denial of the No-Difference view, means that the value of A depends on whether we are comparing it to B or some other outcome. For suppose we are comparing A not only to B, but also to C, as in the following extension of Case 1:[44]

	i	***j***	***k***
A	10	5	Ω
B	Ω	10	5
C	5	Ω	10

Extended Case 1

In this case, all three options are obviously equally good, so B cannot be better than A. This means that to deny the axiological version of the No-Difference View in Case 1, one would have to claim that the value of A and/or B varies between Case 1 and Case 1*. I find that hard to believe. Instead, I agree with Parfit who writes about an analogous case that "[t]he goodness of outcomes could not, I believe, vary in the way described by [this claim]."[45] This is also captured by Anonymity as introduced above. A, B and C are permutations of each other. They are identical in all features that we normally think can contribute to an outcome's value, like total well-being, average well-being and the degree of equality. It is only that the "owners" of the two well-being levels differ. This, however, should not make a difference to our evaluative assessment of them. In other words, I believe that the goodness of an outcome is entirely determined by its internal features. This is what Temkin calls the *Internal Aspects View of Outcome Goodness*,

[44] This case is structurally analogous to one of Parfit's cases. See Parfit, "Future People, the Non-Identity Problem, and Person-Affecting Principles." p. 144.
[45] Parfit. 145.

which I think is correct.[46] I, therefore, do not find this approach of denying the No Difference View on axiological grounds very promising. Let me now move on to Impersonal Value

Impersonal Value

An impersonal approach to axiology does not share the problems of either of the above approaches. On such an approach, it does not matter whether we believe that existing at a positive well-being level is better *for* the respective person than nonexistence, since in either case, the value located in this person's life contributes to the impersonal value of the distribution.

Many theorists, however, reject the notion of such impersonal value. The most common objection is that impersonal theories fail to take persons seriously. Consider (impersonal) utilitarianism, which, in Johann Frick's words

> [b]y treating the moral significance of persons as derivative of their contribution to valuable states of affairs in which total well-being is maximized […] reverses what strikes most of us as the correct order of dependence. Human well-being matters because people matter – not vice versa.[47]

And Ralf Bader writes that

> [o]n this approach persons are treated as mere containers of impersonal good that are dispensable and replaceable. This form of utilitarianism cares only about an abstract and

[46] See Temkin, *Rethinking the Good : Moral Ideals and the Nature of Practical Reasoning*. 370. Note that I am not claiming that Anonymity and the Internal Aspects View amount to the same thing.
[47] Johann Frick, "On the Survival of Humanity," *Canadian Journal of Philosophy* 47, no. 2–3 (May 4, 2017): 344–67. p. 351.

> impersonal form of goodness, not about persons and what is good for them, and consequently does not show adequate concern for persons.[48]

However, even if correct when concerning various forms of utilitarianism, this objection only has force when we employ a moral theory according to which the evaluative dimension exhausts the moral space. In such a theory, if persons are not taken seriously in the evaluative part of the theory, then they are not taken seriously at all.

If, however, person-affecting considerations figure in our moral theory independently from evaluative considerations, then it is no longer the case that we fail to take persons seriously full stop. We just fail to take persons and the boundaries between them seriously in one dimension of our moral theory. Our theory could, however, still have other elements that show proper concern for persons. These elements, however, will not be found in the axiological dimension of our theory. Rather, they will enter our deontic judgements about the choices we face through some other route. So, for the opponent of theories that employ impersonal value, it is not enough to show that these theories fail to take persons seriously in their axiology. For, even if that were the case, they could still take persons seriously in some other way, which is precisely what a pluralist approach to population ethics of the kind I am interested in tries to do.[49]

---

[48] Ralf M. Bader, "Person-Affecting Utilitarianism," in *Oxford Handbook of Population Ethics*, forthcoming. p. 4.

[49] This criticism thus misses its mark much in the same way as criticisms of "good period" in the Moorean sense often do. Such critics of Moore's theory miss their mark because they object to the impersonal construal of axiological value when what they really want to object to is Moore's utilitarianism. See Regan, "Why Am I My Brother's Keeper?"

A second line of criticism concerns the fact that since the value located in this person's life contributes to the impersonal value of a distribution, adding a person with a good life makes a distribution better, other things equal. Therefore, any approach to population ethics that employs an impersonal account of value is bound to deny the Neutrality Intuition. I now move on to discuss this intuition and the closely related *Procreation Asymmetry.*

III. Neutrality

The Neutrality Intuition (henceforth NI) is the intuition that adding a person to a population is in itself an "ethically neutral" act, at least as long as their life is worth living. On the other hand, when a person's life would be worth not living, i.e. she would have negative lifetime well-being, most people think that this gives us a reason not to add this person. In population ethics, this pair of judgements is known as the *Procreation Asymmetry* (Asymmetry for short).

> *The Asymmetry.*
>
> (1) That a person would have a life that is worth not living provides a moral reason not to cause that person to exist.
>
> (2) That a person would have a life worth living does not, on its own, provide a moral reason to cause that person to exist.[50]

[50] The first discussion of the Asymmetry is in Narveson, "Utilitarianism and New Generations." My formulation is very close to Jeff McMahan's. See Jeff McMahan, "Asymmetries in the Morality of Causing People to Exist," in

I will take the first half of the Asymmetry for granted. The challenge posed by the Asymmetry is to defend the second half without giving up on to the first. I am sceptical that this challenge can be met. The second half amounts to a deontic reading of the NI.

> *Neutrality Intuition (deontic)*. That a person would have a life worth living does not, on its own, provide a moral reason to cause that person to exist.

The NI can also be given an axiological interpretation.

> *Neutrality Intuition (axiological).* Adding a person that would have a life worth living to an outcome *A*, does not, on its own, make *A* better.

For now, I focus on the axiological version of the intuition, but will return to the deontic version further down. How does the axiological version of the intuition sit with the different categories of axiologies we considered above? To answer this question, reconsider Case 3.

	i	***j***
A	9	Ω
B	9	10

Case 3

We have seen that a plausible impersonal axiology judges B to be better than A.[51] This amounts to a rejection of the NI. Adding *j* makes the outcome better, simply because she would have a life worth living, thereby maximizing the general value. It does not matter whether B is better *for j* than A.

Harming Future Persons, ed. Melinda A. Roberts and David T. Wasserman, vol. 35 (Dordrecht: Springer Netherlands, 2009), 49–68. p. 49.

[51] Plausible in that it is not extreme or absolutist in its concern for equality.

Things look different under a person-affecting construal of axiology. As we have seen, this depends on our stance towards Comparativism. If we reject Comparativism, then B is not better than A for anyone, and thus not better than A overall. Rather, A and B are incomparable. This kind of theory upholds the NI, but I considered and rejected it above.

That leaves the combination of a person-affecting axiology and Comparativism. I have argued above that the only remotely plausible version of Comparativism is Restricted Comparativism, which would allow us to say that B would be better for *j* than A, without committing us to the absurd claim that A would be worse for *j* than B. Can this view hold on to the NI? The answer is not straightforward. This is because, as we have seen, this position violates Axiological Invariance. If we choose B, then B will be better than A, and it seems that the NI is violated. If we fail to choose B, however, then it is not the case that B is better than A, and the intuition is not violated. On this account then, we will not reach the judgement that B is better than A overall, thus avoiding violation of the letter of the NI.

However, I think it is clear that we still violate the NI spirit. Suppose we face the choice exhibited in Case 3 and look at the reasons in favour of one over the other option. We know the following. If we choose B, this will be better for *j* than had we not chosen B. If we choose A, on the other hand, we will know that had we decided otherwise, this would have been better for *j*. These considerations clearly count in favour of B over A. It is important to note that these are axiological reasons, nonetheless.

Thinking about the issue in this way, makes it hard to see a clear path for Restricted Comparativism to deny the axiological NI. However, this puts the account in good company, as there is an effective argument against the axiological NI that does not depend on any specific construal of axiology. Consider the following three outcomes:[52]

	i	***j***
A	9	Ω
B	10	7
C	9	10

Mere Addition

It seems intuitively clear that the following two pairwise judgements are true.

(1) B is better than A.

(2) C is better than B.

The first claim can be defended in the following way. Everybody who exists in A, is better off in B. This clearly is one way in which B is better than A. The only other difference between the two outcomes, is that in B there is an extra person, *j*, who has a life well worth living. Even if we don't think that this fact makes an outcome better by itself (i.e., the NI), at least it should be clear that it does not make it worse. This claim is captured by the

[52] This case and the following analysis is a variant of Parfit's "Mere Addition Paradox". See Parfit, *Reasons and Persons*. chapter 19.

> *Mere Addition Principle*. Let B be a state of affairs that is just like A except that, in addition, some extra people with lives worth living exist in B who do not exist in A. Then B is not worse than A.[53]

Of course, we could still deny that B is better than A, and instead claim that they are incomparable. As we have seen above, however, this stance would commit us to intuitively highly implausible implications.[54] We should therefore hold that B is better than A. In other words, we should adhere to the

> *Benign Addition Principle*. Let B be a state of affairs that is just like A except that, everyone who exists in A is better off in B, and some extra people with lives worth living exist in B who do not exist in A. Then B is better than A.

I think we should accept this principle. This grounds the first of the above two claims. The second claim can be defended in the following way. Although, admittedly, C is worse than B for *i*, it seems better in all other respects. First of all, it is better for *j*, and to a greater extent than B is better than C for *i*. The benefit we could give *j* by moving from B to C, is greater than the benefit we could give *i* by moving from C to B. Second, C is better than B in all impersonal, or anonymous, respects. Total and average well-being are greater in C and, moreover, well-being is also more equally distributed in C than it is in B. We should therefore hold that C is better than B overall.

---

53 This is Hilary Greaves' formulation. See Greaves, "Population Axiology." p. 3.

54 See my discussion of Bader's person-affecting theory above.

From these two claims and the transitivity of "better than," it follows that C is better than A. We thus have a case in which merely adding a person with a life worth living makes an outcome better, i.e. C is better than A and the axiological NI is violated.[55]

So far, I have shown that on each construal of axiology that we considered, there is no clear path to holding on to the intuition, at least not without accepting widespread incompleteness. In addition, we have seen an argument that does not rely on any specific construal of axiology, but only on very modest and intuitively attractive premises. This overdetermines the case against the axiological NI. What then about the deontic construal of the NI? Recall

> *Neutrality Intuition (deontic)*. That a person would have a life worth living does not, on its own, provide a moral reason to cause that person to exist.

First of all, I should note that I believe that one of the main reasons why many people are attracted to the axiological construal of the intuition is that they mistake it for this deontic one. They, for example, might believe things like "no one is under a moral obligation to have children" or "there don't seem to be any moral reasons for having as many children as possible." Such reasoning, however, is too quick. For it is not at all obvious that such claims would follow from a rejection

---

[55] Of course, we could also attempt to deny the transitivity of "better than." See Temkin, "Intransitivity and the Mere Addition Paradox." I shall not consider this very revisionary possibility here. For a slightly less revisionary solution to the Mere Addition Paradox along somewhat similar lines, see Frick, "Context-Dependent Betterness and the Mere Addition Paradox." Frick argues that while the "better than" relation is transitive, it might be "context-dependent." I am sceptical about this approach, but do not have the space to do it justice here. In brief: while I do think that "moral choiceworthiness" can be choice set dependent, "goodness" cannot.

of the axiological NI. This depends on the moral theory that connects our axiology to moral obligations. In particular, if we had a moral theory that did not include any axiological reasons, there would be no reason at all to think that any of these implications would follow from the axiological NI.

Furthermore, bringing people into existence has drastic effects on the lives of other people. In practice, it is *never* the case that an additional life has *no* effects on other people. If, on the other hand, we could push a button and happy people with fantastic lives would pop up in a galaxy far away from ours, the case for the axiological NI looks much weaker. In effect, however, it is such a scenario we should be thinking about when discussing the *principled* reasons for and against the axiological NI. We should think of cases in which no other factors are at play.

In such a case, the rejection of the deontic NI *does* follow from the rejection of the axiological NI, if we assume a moral theory that takes into account axiological reasons. This is because, on any such a theory, the fact that one outcome X is better than another outcome Y provides us with a reason to bring about X rather than Y. The pluralist account I am proposing in this essay would give rise to a theory of this kind. Moreover, not only would we have a reason to bring into existence lives worth living. There would also be cases in which we would have an obligation to bring these lives into existence. This follows from the following principle:

> *From Reasons to Obligation.* If you have a reason to choose X over Y, and no reason to choose Y over X, then you have an obligation to choose X over Y.[56]

To me, this principle seems hard to reject. It is a very weak principle, since it allows that *any* reason to choose Y, moral or otherwise, is enough to block an obligation to choose X. But, if there really is *no reason at all* to choose Y, and there is a reason to choose X, then I don't see how failing to choose X could be justified.

I should note that some might be inclined to deny this principle in the domain of prudence. For example, if you have never tried a mangosteen, you *ought* to do so, since mangosteens are delicious. However, it might be a bit of a stretch to say that you have an *obligation* to try the mangosteen even if you have no countervailing reason not to.[57] I can see that, intuitively, it might seem odd that we have a moral obligation to promote our own good. However, I agree with Ewing, who writes the following:

---

[56] For a general principle of this sort, see e.g. John Broome, "Normative Requirements," *Ratio* 12, no. 4 (December 1999): 398–419. p. 400, who phrases it in terms of "ought" rather than "obligation."

[57] This is John Broome's example, though I do not know whether he would agree with the way I apply it here. See John Broome, "Reasons," in *Reason and Value : Themes from the Moral Philosophy of Joseph Raz*, ed. Philip Pettit et al. (New York: Oxford University Press, 2004). p. 40. Similarly, providing another fruit-related example, Véronique Munoz-Dardé writes that "[a]fter all, it is not immediately obvious that someone who feels the attraction of eating a peach, but in the end fails to do so, though no other consideration intervenes in their thoughts, thereby exhibits practical irrationality." See Véronique Munoz-Dardé, "The Distribution of Numbers and the Comprehensiveness of Reasons," *Proceedings of the Aristotelian Society* 105 (2005): 191–217. at p. 203. See also Michael Stocker, "Agent and Other: Against Ethical Universalism," *Australasian Journal of Philosophy* 54, no. 3 (December 1976): 206–20.

> If I had not decided that it was a duty ceteris paribus to further one's own pleasure, I should have been forced to deny either that my own pleasure is a good or that it is a moral duty to pursue the good as far as we can. To my mind this is a strong argument for my conclusion about [an obligation to further one’s own] pleasure, since I have a strong inclination to think both these propositions self-evidently true.[58]

Going forward, I will assume the principle From Reasons to Obligation to be correct. In terms of other-regarding reasons, which I am mostly concerned with in this essay, this is not a very controversial assumption to make anyways.

Let us take stock. One of the assumptions we started out with was that in order to account for our intuitions in some variable population cases, like Case 2, our moral theory needs to take into account axiological reasons stemming from the overall value of outcomes. I have considered various ways of cashing out this notion of “general value” and have argued that the most promising way to do so is in terms of impersonal value. This straightforwardly leads to a rejection of the axiological interpretation of the NI. However, even if we reject impersonal axiology and insist on (the most plausible version of) a strictly person-affecting axiology, we are nonetheless led to reject the axiological NI, or so I have argued. Via the assumption that some of our reasons stem from our axiology, this in turn, should lead us to also reject the deontic NI. That is, we *do* have (axiology based) reasons to bring people into existence merely because their lives would be worth living.

Going forward, I will thus assume that axiology should be impersonal in the above sense. For reasons of simplicity, I shall further assume that

[58] Ewing, *The Definition of Good*. p. 162.

this axiology takes the total utilitarian form. However, nothing hinges on this assumption.

I have claimed at the outset that axiological reasons, however, cannot account for our intuitions in cases like Case 1. In addition, we also need to take into account other moral reasons, namely reasons stemming from the complaints that people could raise against our choices. Let us now examine these sorts of reasons.

IV. Whose Complaints?

Consider Case 1 again.

	i	j	k
A	10	5	Ω
B	Ω	10	5

Case 1

I have claimed that in this case, we ought to choose B although the two outcomes are permutations of each other and are thus equally good. I said that the reason why we ought to choose B is that if we failed to do so, *j* could raise a justified complaint against our choice. We could have easily made her much better off but failed to do so. This suggests the following.

> *Grounds for Complaint*: If you bring about outcome X, an individual *i* can raise a complaint against your choice iff *i* exists in X and there is an alternative outcome Y, in which *i* would have been better off.

Such complaints are reasons to choose Y over X. One question that immediately poses itself is the following. How much weight do these reasons carry vis à vis other reasons, in particular, axiological reasons? If in Case 1, we ought to choose B, we know that they carry *some* weight. But given that there is an "axiological tie" between A and B, this weight might be very small. It could just be that they might serve as tie-breakers, but no more. In order for them to have any real bite, we must be prepared to pay some axiological cost. In other words, there is some ε, such that we still ought to choose B in the following case.

	i	***j***	***k***
A	10+ε	5	Ω
B	Ω	10	5

Case 1+

But, on the other hand, there is also some larger ε, such that we ought to reverse our choice and choose A. If the axiological costs become too high, the complaint-based reasons are no longer weighty enough. Of course, finding the right weight to put on these different categories is a formidable challenge for any pluralist theory and shall have to say a little more about it at the very end of this essay. For now, let me just say that for the theory to be meaningful, there must be some axiological cost we are willing to accept to heed complaints, but there also must be some limit on the cost we are willing to accept. For the remainder of this essay, however, I will assume that ε is such that we should still choose B in this case.

Having mapped out the general structure of the approach I am proposing, I now move on to one such account that has recently been proposed by Michael Otsuka.

Otsuka's Account

Here is Otsuka's proposal:

> *An Existential-Choice-Tracking Proposal:*
>
> (i) When someone's existence is choice-independent, her complaints are morally weighty. They are sufficiently weighty that they are often capable of overriding anonymous moral considerations that tell in the other direction.
>
> (ii) When, by contrast, someone's existence is choice-dependent, her complaints retain moral weight. But this weight is greatly reduced in comparison. Such complaints are [almost] never sufficient to override anonymous considerations that pull in another direction.[59]

What Otsuka here calls "anonymous moral considerations" refer to what I above called the impersonal value of an outcome.[60] So, according to Otsuka, when individual complaints against an impersonally better distribution are very strong, then we sometimes need to do what heeds these complaints, even if we thereby choose a worse outcome. This, of course, is the same general idea that I have been proposing.

Otsuka draws a principled distinction between complaints by people whose existence is choice-independent (call these "independent

---

[59] Otsuka, "How It Makes a Moral Difference That One Is Worse off than One Could Have Been." p. 199. In reply to one objection, in the end Otsuka weakens his principle with the "almost" clause in (ii). This need not concern us here.

[60] Otsuka objects to the label "impersonal" value or goodness. This is, because "such goodness paradigmatically consists of people's lives going well – their flourishing, the satisfaction of their needs, their freedom from suffering, among others", and so thinks it "misleading to describe it as impersonal." See Otsuka. n. 4. I think Otsuka has a point. However, for purposes of keeping in line with the literature's terminology, I shall continue to use the label "impersonal."

people") and complaints by people whose existence is dependent on our choice (call these "dependent people"). To see the importance of this distinction for Otsuka's account, consider the following case, which is a variation of Case 1+.

	i	***j***	***k***
A	10+ε	5	Ω
B	Ω	10	5
C	Ω	Ω	Ω

Case 6

Remember that in Case 1+, for a small enough ε, my complaints-based approach had us choose B. Otsuka would agree with this judgement.[61] In a case like Case 6, however, he argues that B is impermissible and we should choose either A or C.[62] For Otsuka, C "is a morally permissible option, on grounds that one is not obliged to bring more people into the world who will have lives worth living."[63] We can thus permissibly choose C. We may, however, not choose B.[64] Otsuka would argue that unlike in Case 1, *j's* complaint against A here is not sufficient to outweigh the loss in overall goodness compared to B. According to Otsuka, this is because "if [A] is chosen, there will be

[61] See Otsuka. pp. 192-193. Note that in Otsuka's cases, well-being is represented by the number of years a person lives in a specific outcome, which he takes to correspond "to the interpersonally comparable cardinal level of welfare that this child enjoys over her lifetime." This is a mere representational difference.
[62] This choice seems to violate Contraction Consistency. According to this principle, if an option is (permissibly) chosen out of a set, then it is also chosen out of a subset containing this alternative. Otsuka is aware of that problem and argues, convincingly in my view, that this principle is either false or best understood in such a way that it is not violated by his view. See Otsuka. p. 204 and section VI below.
[63] Otsuka. p. 203.
[64] Remember, I am assuming ε to be such that we should still choose B in Case 1+.

another option that is worse for [*i*], namely, [C], in which [*i*] does not exist."[65] He goes on to say that

> [i]n contrast to Case 1, the choice of [A] in [Case 6] does not condemn [*j*] to her worst option. Rather, [A] is much *better* for [j] than an alternative [C*]* that is a genuine moral option for you. The benefit to [*j*] of your choice of [A], in comparison with [C], is [5]. The fact that [*j*] would already receive such a benefit […] via your choice of [A] in comparison with [C], weakens her case for a second helping […] via your choice of [B] instead of [A].[66]

From this discussion, it emerges that Otsuka accepts both a form of Comparativism and (at least the weak version of) the deontic NI, according to which there can be no *obligation* to bring people into existence merely because they would have lives worth living.

The weak deontic NI is widespread, which is why, I take it, Otsuka doesn't provide us with an argument for it. However, we have seen in the preceding sections that such an argument is needed. It is very difficult to hold onto any form of the NI. This, in particular, is the case if one accepts the reason giving force of impersonal value, as Otsuka does. It would require him to either defend the axiological NI or reject the principle I called From Reasons to Obligation. The first option is all but a rejection of the notion of impersonal (or anonymous) value itself, while the second would require a revision of common-sense normative thinking.

So, avoiding this latter path, Otsuka could reject the notion of impersonal value altogether. This, however, would require him to take on board the difficulties of person-affecting axiology that I pointed out

[65] Otsuka, "How It Makes a Moral Difference That One Is Worse off than One Could Have Been." p. 201.
[66] Otsuka. p. 203.

in section 2. However, even then, as we have seen in the preceding section, it is still unclear how he could resist the case against the axiological NI without accepting widespread axiological incompleteness.

Could Otsuka instead give up on the NI? He would then have to acknowledge that if a person would have a life worth living and her existence would have no effects on other people's lives, then we do have an obligation to bring her into existence. Now, even if Otsuka was prepared to accept this much, there is a problem for his account. Consider the following case:

	i	***j***	***k***	***l***	***m***
A	10+ε	5	Ω	Ω	Ω
B	Ω	10	5	Ω	Ω
C	Ω	Ω	Ω	4	4

Case 7

Otsuka writes that in a case like this, we ought to reason as in Case 1+ and choose B.[67] On first reflection this seems odd, given that in Case 6, which is just like this case without *l* and *m*, Otsuka had us choose A. Here is Case 6 again, this time with two extra columns for *l* and *m*.

	i	***j***	***k***	***l***	***m***
A	10+ε	5	Ω	Ω	Ω
B	Ω	10	5	Ω	Ω
C	Ω	Ω	Ω	Ω	Ω

Extended Case 6

[67] See Otsuka. p. 200.

Otsuka's reason for treating Extended Case 6 and Case 7 differently, would be that unlike C in Extended Case 6 (the empty population), he thinks that C in Case 7 "is not a genuine moral option even if it is a possible object of choice."[68] Such options should not figure in our moral deliberation and the addition of C should not change our assessment of the case. This does sound appealing as there seems to be no reason at all to choose C in Case 7, given the availability of options A and B. According to this proposal, we should thus exclude C in Case 7 at the outset. This would allow Otsuka to treat this case just like Case 1+, where *j's* complaint against A compels us to choose B. Had we not excluded C at the outset, the presence of this option would have made it the case that *i's* complaint against A loses its force, because then there would be a "genuine moral option" which would be worse for *j*.

I want to note two problems with this approach. The first is that the notion of "genuine moral option" is not well defined. By deciding whether an option is a genuine one, however, we decide ourselves which people count as dependent and which as independent, which in turn determines the weight of their complaints. This allows us as moral decision makers to shift the parameters that are supposed to serve as fixed points within the theory.

Even if we accept that move, however, the second problem for Otsuka is the following. If we reject the NI and acknowledge that we ought to bring people into existence if they have lives worth living and don't affect the lives of other people, then in the following case we ought to choose B:

---

[68] Otsuka. p. 200.

	i	*j*	*k*	*l*	*m*
A	Ω	Ω	Ω	Ω	Ω
B	Ω	Ω	Ω	4	4

Case 8

Given this, it seems very odd to say that C in Case 7 is not a genuine moral option and should be excluded at the outset, but if we replace it with the empty population (and thus transform Case 7 into Extended Case 6), we should not exclude C. This amounts to claiming that an option (B) that should be chosen over a genuine moral option (A) in a pairwise choice, is not itself a genuine moral option in the same choice context in which A *would be* a genuine moral option. I fail to see how a moral theory could have this structure. This structure, however, follows for Otsuka's account from rejecting the NI.

I shall now suggest that there is another way of formulating a pluralist account that does not run into the same problems as Otsuka's. For the remainder of this essay, I will be concerned with formulating and defending this account.

A Different Pluralist Account

My first step is to reject all versions of the Neutrality Intuition. This follows from my accepting the impersonal construal of general value, the reason giving force of this general value, and the principle From Reason to Obligation. Next, I reject Comparativism. This saves us from metaphysical obscurities and, once we accept impersonal value, we do not have much need for being able to compare the personal value of existence with nonexistence, anyways.

We have seen in sections 2 and 3 that such an axiology has the best shot at coherence without giving rise to widespread incompleteness. It follows from this that we should concede that there is no moral relevance to the distinction between independent and dependent people. Instead, let me introduce a different distinction that, so far, I have not made explicit. This is the distinction between what we might call *uniquely realizable* and *non-uniquely realizable* people. A uniquely realizable person is one whose life has the same well-being level in all outcomes in which she exists. A non-uniquely realizable person is one for whom this does not hold. Such people have at least two possible well-being levels that we could realize for them.

It is important to note that this is a different distinction from the one between *independent* and *dependent* people. These two distinctions cut across each other. We can see that in the following case.

	i	***j***	***k***	***l***
A	5	5	5	5
B	Ω	10	5	10
C	Ω	Ω	5	10

Case 9

In this case, *i* is a *dependent and uniquely realizable* person, *j* is a *dependent and non-uniquely realizable* person, *k* is an *independent and uniquely realizable* person, *l* is an *independent and non-uniquely* realizable person. All non-uniquely realizable people can raise a complaint when we fail to make them as well off as we could have by bringing about an outcome in which they fare worse than in another outcome that was available to us. Uniquely realizable people on the other hand can never raise a complaint since the notion of a complaint as defined so far only makes sense when there was another outcome available in which they would have been better off, which can't be the case for uniquely realizable people.

If we reject Comparativism, as I do, then it does not make sense to say that the complaints of dependent people carry less weight than those of independent people. Rather, the important distinction now lies between uniquely and non-uniquely realizable people. This is because only non-uniquely realizable people can ever raise a complaint. This explains why in Case 1+ we ought to choose B. Here is the case again:

	i	***j***	***k***
A	10+ε	5	Ω
B	Ω	10	5

Case 1+

We ought to choose B, because *j* can raise a complaint against A. In this case, the distinction between dependent and independent people coincides with the distinction between uniquely and non-uniquely realizable people. *j* is an independent, non-uniquely realizable person, while *i* and *k* are dependent, uniquely realizable people. The distinctions come apart, however, in Case 6.

	i	***j***	***k***
A	10+ε	5	Ω
B	Ω	10	5
C	Ω	Ω	Ω

Case 6

Here, all of *i*, *j* and *k* are dependent people. *j*, however, still is non-uniquely realizable and if we choose A, his complaint will carry just as much weight as his complaint if we choose A in Case 1. I believe that in Case 6 we have an obligation to choose B. We have an obligation not to choose C, because we would thereby gratuitously avoid realizing a much better outcome in terms of impersonal value. We can thus rule out C. Out of the remaining A and B, even though A is a slightly better outcome than B, we ought to honour the complaint *j* could raise against A and choose B.

On this approach, then we always ought to pick the best outcome, unless one or more people can raise a significant complaint against this outcome. Such complaints can only be raised by non-uniquely realizable people. The complaints of dependent and independent people carry the same strength. This strength turns on how much better they would fare in other available outcomes. In other words, sometimes very strong individual complaints can block the realization of what would be otherwise required. So, I reject the claim that we always ought to do what brings about the best outcome. Sometimes we ought to bring about an all-things-considered inferior outcome to honour the complaints of specific people. Unlike Otsuka, however, I reject the claim that we can never have an obligation to bring people into existence simply because their lives would be worth living. This follows from the fact that, absent strong opposing individual claims, we have an obligation to realize the best outcome. Since new lives add value to an outcome, we sometimes have an obligation to create them.

So much for the general exposition of the pluralist account of population ethics that I want to propose in this essay. I now need to lift one limiting restriction to all of the examples we have dealt with so far. This is the restriction that we only consider lives that are worth living.

V. Negative Well-Being Levels and the Weak Asymmetry

Any approach to population ethics needs to be able to take into account lives with negative well-being levels, i.e. lives that are "worth not living." So far, however, I have only dealt with cases in which all existing people have positive well-being levels. Once we introduce negative well-being levels, our intuitions often radically change. We might want to say, for example, that a person that is created at a negative well-being level when the only alternative was non-existence, *does* have grounds for complaint. If we reject Comparativism, however, this would not be possible on my account. Recall

> *Grounds for Complaint*: If you bring about outcome X, an individual *i* can raise a complaint against your choice iff *i* exists in X and there is an alternative outcome Y, in which *i* would have been better off.

To see that this might spell trouble consider the following case.

	i	***j***
A	10	-9
B	Ω	Ω

Case 10

In terms of impersonal value, A is better than B, so axiology-based reasons speak in favour of choosing A.[69] Still, I believe that many will be very much drawn to the claim that we ought to choose B. If we choose A, *j* would lead a life worth not living, a life of suffering. In order to avoid this, many will be prepared to forego the slight

[69] As before, I am assuming that no facts other than the sum of well-being contribute to an outcome's general value. E.g. the inequality in A does not subtract from A's value.

advantage in terms of general value that A has over B. This, however, cannot be explained by the complaints-based reasoning we have been employing until now. This is because, according to my definition of Grounds for Complaint, *j* would not be in a position to raise a complaint against our choice of A, since there was no other outcome available in which he would have been better off.

Still, it seems natural to say that *j* could very well raise a justified complaint against our choice, given that we caused him to live a life full of suffering. I agree. We need to qualify the above claim that only non-uniquely realizable people can ever raise complaints. This speaks in favour of expanding the grounds for complaint to the following.

> *Grounds for Complaint**: If you bring about outcome X, an individual *i* can raise a complaint against your choice iff *i* exists in X and there is an alternative outcome Y, in which *i* would have been better off, *or i* has a negative well-being level in X and would not have had a negative well-being level in Y.

This to me seems like a natural extension of the complaints-based reasoning. When we choose A in the above case, although *j* would not be in a position to point to an alternative outcome in which he would be better off, he could point to an outcome in which he would not be suffering. This fact seems to give rise to a complaint on his part. Note that this is not the same as saying that B would be *better for* him. As we have seen in the above discussion of Comparativism, B would not be better for him, since he would not exist.

I now want to suggest that this expansion of the complaints-based part of the pluralist approach to population ethics makes it possible to account for a weak version of the Asymmetry.

The Weak Asymmetry.

(1) That a person would have a life that is worth not living provides an axiological as well as a complaints-based reason not to cause that person to exist.

(2) That a person would have a life worth living provides only an axiological reason to cause that person to exist.

This weaker version does not get those attracted to the original Asymmetry all that they want, but it provides the start of a justification for the intuition that we have a special reason not to cause miserable people to exist that goes beyond those axiological reasons that stem from the general value of outcomes.

VI. Objections

The Problem of High Axiological Stakes

So far, I have said little about the relative importance of reasons stemming from complaints vis á vis reasons stemming from the general value of outcomes. I have only dealt with cases where things looked clear, intuitively. These are cases, in which we choose between two alternatives that are (almost) equal in value. In these cases, complaints were able to decide things in favour of one of the options, effectively acting as tie-breakers. Things will look less clear in cases in which considerations of general value speak heavily in favour of one option and complaints in favour of another option. As I briefly noted above, this sort of calibration of the weights we put on different sorts of reasons is a general problem for any pluralist approach. Still, I don't want to shy away from addressing it here. This is partly because the context of population ethics poses problems for a pluralist approach that are different from the usual.

Likely, we are just at the very beginning of human history. This means that almost all the potential good of humanity lies in the far future. What does this fact imply regarding our thinking about population ethics? For one thing, it implies that any theory in population ethics that takes into account axiological reasons needs to be partly oriented towards the very far future. It is not enough to take into account only the easily foreseeable effects of one's action.

If we are straightforward maximizing consequentialists and only take into account axiological reasons, then this observation would imply that what we ought to do depends almost entirely on the way in which our actions could influence the far future. We would have to basically disregard the more immediate effects of our choices. This will strike most people as implausible. It is certainly hard to square with common-sense morality. Nonetheless, the conclusion forces itself upon us once we realize just how early we probably are in human history, and once we accept that some of our moral reasons are axiological. Since, in light of this vast potential value, on any theory that takes into account axiological reasons at all, we are under pressure to conclude that these reasons will wash out any other considerations.[70] Even if

[70] For a detailed account in support of this claim, see Hilary Greaves and William MacAskill, "The Case for Strong Longtermism," *GPI Working Paper 7*, 2019. For a reply, see Andreas L Mogensen, "Staking Our Future: Deontic Longtermism and the Non Identity Problem," *GPI Working Paper 9*, 2019, 32.

axiology were just a tiny part of our theory, if the axiological reasons become weighty enough, this part of our theory will almost entirely determine what we ought to do.

This observation does not change the structure of my proposal, but it does point out the fact that it could be toothless in practice. In fact, the practical implications of a pluralist proposal like mine would come very close to those of the underlying axiology when coupled with consequentialism. For the purposes of this paper, I have often assumed the total utilitarian axiology. This would imply that in many cases my proposal would require us to act *as if* we were utilitarians. This by itself would be an interesting conclusion and I do not (yet) have much to say in terms of trying to resist it. Maybe it is correct. However, in any case, it is not an objection against the *structure* of my proposed account.

The Problem of Improvable Life Avoidance

The objection that I want to consider next is that there might be cases in which my approach could recommend not bringing someone into existence at all, *because* they would have a complaint if we did. In other words, on this approach, we might be led to avoid bringing into existence someone who would have a life that could be better. Hence the term "problem of improvable life avoidance." [71] To see this problem, consider the following case.[72]

[71] I borrow this term from Jacob Ross, "Rethinking the Person-Affecting Principle," *Journal of Moral Philosophy* 12, no. 4 (August 11, 2015): 428–61. p. 443.

[72] This is an adapted version of a case by Ross. See Ross. p. 444.

	i	*j*	*k*	*l*
A	10	5	Ω	Ω
B	9	9	Ω	Ω
C	Ω	Ω	9-ε	9-ε

Case 11

In this case, in terms of general value, B is better than both A and C. However, one could argue that the complaints-based approach tells against choosing B, since this would give rise to a complaint on *i's* behalf, and we should choose C to avoid this complaint. The objection now holds that it is unreasonable to choose C in a situation in which B was also available.

I think there are two ways to answer this objection. The first is to bite the bullet and claim that it is not unreasonable to choose C in this case. If we take complaints seriously, we must be willing to pay some price to avoid them, as I have already argued above.[73]

However, even if we are not willing to bite the bullet in the way I described, there is another reply available. We could argue that though choosing B would give rise to a complaint on *i*'s behalf, *i* would not be justified in pressing this complaint. After all, the only way to avoid his complaint, while still bringing him into existence, would be to choose A. This, however, would give rise to a much stronger complaint on *j*'s behalf. Thus, we could argue that choosing B is justifiable to *i*. In Scanlonian contractualism this is given by what has become known as

[73] See also Otsuka's discussion in light of a similar problem, in Otsuka, "How It Makes a Moral Difference That One Is Worse off than One Could Have Been." pp. 206-9.

> *The Greater Burden Principle.* [I]t would be unreasonable […] to reject a principle because it imposed a burden on you when every other alternative would impose much greater burdens on others.[74]

Unfortunately, although this latter reply is available in this case, it does not help with all cases that give rise to the problem of improvable life avoidance. Take, for example, the following case:[75]

	i	***j***	***k***
A	9	5	Ω
B	Ω	5	8
C	10	Ω	Ω

Case 12

In this case, the problem of improvable life avoidance resurfaces in a slightly different way. Though A gives rise to the best outcome, it also gives rise to a complaint on *i*'s behalf. After all, we could also have chosen C, in which *i* would have been better off.

In this respect, the two cases are similar. Unlike A in Case 11, however, C in Case 12 does not give rise to a complaint itself. So, the second reply we gave to the problem in Case 11, is not available in this case. It is not the case that we can claim that *i* would be unjustified in pressing his complaint against A. Therefore, *i*'s complaint against A stands. To avoid this complaint, we could choose C or B, neither of which would itself give rise to a complaint. Since C is much worse than B in terms of general value, it might be that we ought to choose B, thus avoiding bringing *i* into existence at all. In a way, this would seem to make *i* a

[74] T.M. Scanlon, "Contractualism and Utilitarianism," in *Utilitarianism and Beyond*, ed. Amartya Sen and Bernard Williams (Cambridge University Press, 1982), 103–29. at p. 111.
[75] I thank Johann Frick for raising a similar case to me in discussion.

"victim of his own complaint."[76] Rather than bringing *i* into existence in a state in which he would have a life well worth living, but could be better off still, had we chosen otherwise, we choose not to bring him into existence at all. Put differently, by giving rise to a complaint, the availability of an option, C, which we would not have chosen anyways, affects our choice between two other options, A and B. How problematic is this implication?

To answer this question, it is useful to uncover the structural underpinnings of the problems of this section. I claim that what gives rise to these at first blush counterintuitive implications, is the fact that our theory violates the following principle.

> *Contraction Consistency*. If an alternative is permissibly chosen from of a set of alternatives, then it is also permissibly chosen out of any subset containing that alternative.

Contraction Consistency requires that if alternative X is permissibly chosen out of the set {X,Y,Z}, then X must also be permissibly chosen out of the set {X,Y}. Often this seems to be a basic requirement of rational choice. Suppose you are trying to decide between eating chocolate, banana or strawberry ice cream. You have settled on chocolate just before the waiter tells you they are out of banana. You then think, "oh, in that case I'll have strawberry." Such reasoning seems profoundly flawed. In most cases your decisions should not depend on the availability or absence of an option you wouldn't

[76] I thank Frick for this expression, which, of course, is not entirely correct, since *i* does not exist in B and thus cannot *be* a victim, or anything else for that matter. Still, it is nice way of putting the problem.

choose anyways. Contraction Consistency is undoubtedly very plausible as a desired property of rational choice.

To see that the complaints-based element in our theory gives rise to a violation of Contraction Consistency, consider the following contraction of Case 12 above.

	i	***j***	***k***
A	9	5	Ω
B	Ω	5	8

Contracted Case 12

If A and B are the only available options, it is clear that on the approach we have been considering in this essay, we ought to choose A. In effect, this is just a milder version of my Case 2. A's general value exceeds B's, and since no other reasons are at play, this means that we ought to choose A.

If we now expand this case to the original Case 12 by adding option C, we are introducing a complaint. According to our theory, this gives rise to a moral reason that was not there before. This reason speaks in favour of B, or rather against A. It changes the normative situation in such a way that A is no longer permissible. However, it is also not the case that we now ought to choose the newly introduced C. Rather, we now ought to choose B, which was there all along. This is a violation of *Expansion Consistency*, which is (something like) the flipside of Contraction Consistency.[77] If we reason the other way around, contracting the choice set {A,B,C} to {A,B}, we get a violation of Contraction Consistency.

[77] See Amartya Sen, "Internal Consistency of Choice," *Econometrica* 61, no. 3 (1993): 495–521. Contraction Consistency is Sen's condition α, Expansion Consistency is his condition β.

How problematic is this violation of Contraction Consistency? There are broadly two ways to defend theories against the charge of violating this principle. First, we could argue that Contraction Consistency is actually *not* necessarily required by rationality.[78] Second, we could argue that many choice patterns that seem to violate Contraction Consistency are actually ill-described and better descriptions will show that they actually do not violate it.[79] Though I have sympathies for both approaches, I here focus on the latter.

We can distinguish between the intrinsic and the extrinsic properties of a good. These together make up the overall choiceworthiness of that good. Unlike the extrinsic properties, the intrinsic properties of a good stay the same across all comparisons. Take the following example by Amartya Sen.[80] Imagine three slices of the same cake, small, medium, and large. Suppose that when offered cake, etiquette demands never to choose the largest slice. This means that when offered a choice between all three slices, you may not choose the large slice, but you may choose the medium slice. When, however, offered a choice between only the small and the medium slice, choosing the medium slice is now impermissible and you ought to choose the small slice. Formally this is a violation of Contraction Consistency. Out of the set *{small,medium,large}* we chose *medium* and out of the subset *{small,medium}* we chose *small.* The intrinsic properties of the medium-sized slice, like its taste or its size, have not changed. The extrinsic properties, however, have changed. One of these extrinsic properties

[78] See e.g. Paul Anand, *Foundations of Rational Choice under Risk* (Clarendon Press, 1993). and Bertil Tungodden and Peter Vallentyne, "On the Possibility of Paretian Egalitarianism," *The Journal of Philosophy* 102, no. 3 (2005): 126–154.
[79] For both these options, also see Otsuka, "How It Makes a Moral Difference That One Is Worse off than One Could Have Been." p. 204.
[80] See Sen, "Internal Consistency of Choice." p. 501.

is the property of being the largest slice on offer. *Medium* has this property in the set *{small,medium}*, but not in the set *{small,medium,large}*. From this we can infer that the permissibility of choosing an option can change, even if its intrinsic properties stay the same.

But we need a principled account of when it is rational to distinguish between options based on the available alternatives and when it is not. For it can't be permissible all the time for two reasons. The first is that there are examples, like the ice cream one, where it clearly seems irrational to make your choice depend on what alternatives are available. And the second reason is that, if this choice rationalizing approach of redescribing our choice options were always permissible, then Contraction Consistency and other requirements of rational choice would seem empty. We could simply always redescribe our options in such a way that it became impossible not to conform to those requirements. Rational choice theory would then become pointless as a normative theory. We, therefore, need a principle that tells us when it is permissible to distinguish between options or outcomes, which in our case are distributions, solely based on what other alternatives are available. One such principle is due to John Broome, who says that according to the

> *Principle of Individuation by Justifiers*. Outcomes should be distinguished as different iff they differ in a way that makes it rational to have a preference between them.[81]

Coming back to our case, we need to ask if it is rational to have a preference between A when only A and B are available, and A when

[81] Broome, *Weighing Lives*. p. 103.

A, B and C are available. I think it is. Choosing A when we could have chosen only B instead is not wronging *i* in the way that choosing A when C was also available does. Choosing A in the latter case, gives rise to a complaint on *i*'s behalf that choosing A in the restricted choice between A and B does not. It therefore seems rational to have a preference between the two, because one includes wronging someone in a way that the other does not. And if it is rational to have a preference between the two "individuated" alternatives, then the two can be sensibly distinguished and we don't actually violate Contraction Consistency.

Conclusion

In this essay, I have argued for a pluralist approach to population ethics that combines axiological reasons with a particular kind of non-axiological reasons, personal complaints. On such an approach, when choosing between different distributions of well-being, we need to take account of the overall or general value of each distribution. I claimed that there are broadly two ways to make sense of this notion of general value, person-affecting and impersonal. I have argued that the impersonal approach is more plausible. This is because person-affecting population axiologies either, first, have metaphysically troubling implications regarding the comparison of existence and nonexistence, second, give rise to implausibly widespread axiological incompleteness, or, third, give rise to an implausible invariant structure.

Impersonal population axiologies on the other hand do not have these implications. They have been criticized, however, for failing to take persons seriously. I have argued that even if this objection is warranted, it could be met outside of population axiology, by taking into account non-axiological moral reasons that do take persons seriously. One way to do so is to take into account the personal complaints individual people could raise against our choices. These complaints can sometimes block the realization of what would otherwise be required by the underlying axiology. In other words, the account rejects the claim that we always ought to do what brings about the best outcome. Rather, sometimes we ought to bring about an all-things considered inferior outcome in order to honour the complaints of individual people.

However, if such complaints are not present or not strong enough, I have argued that accepting this reason giving force of the impersonal value of a distribution in some cases gives rise to an obligation to create people merely because their lives would be good. This requires abandoning the popular Neutrality Intuition, according to which bringing people who would have a life worth living into existence, is itself morally neutral. I have shown how this implies that the account also requires giving up on the Population Asymmetry. However, I have shown that it can salvage a weaker version of the Asymmetry. According to this Weak Asymmetry, although we do have reasons to create people because their lives would be worth living, these reasons are merely axiological. On the other hand, we have axiological as well as non-axiological reasons not to create people who would have lives that are worth not living.

Finally, I have replied to two possible objections against the kind of pluralist approach to population ethics I have proposed in this essay. One of these objections, I called the "problem of high axiological stakes." This is the worry that if the axiological reasons become weighty enough, this part of our theory will almost entirely determine what we ought to do. Given that in the context of population ethics, the axiological stakes tend to be very high, it might be that, in practice, we can often ignore non-axiological reasons. In other words, a theory of population ethics that gives *some* weight to axiological reasons might in practice turn out to look like a theory that gives *all* the weight to these reasons. This by itself would be an interesting conclusion. Is it correct? I must leave examining this question to another occasion.

Bibliography

Adler, Matthew D. *Well-Being and Fair Distribution*. Oxford University Press, 2012.

Adler, Matthew D, and Chris William Sanchirico. “Inequality and Uncertainty: Theory and Legal Applications.” *University of Pennsylvania Law Review* 155, no. 2 (2006): 279–377.

Anand, Paul. *Foundations of Rational Choice under Risk*. Clarendon Press, 1993.

Anscombe, G. E. M. “Who Is Wronged? Philippa Foot on Double Effect: One Point.” In *Human Life, Action and Ethics: Essays*, 249–51. Exeter, UK: Imprint Academic, 2005.

Arrhenius, Gustaf. “Can the Person Affecting Restriction Solve the Problems in Population Ethics?” In *Harming Future Persons*, edited by David Wasserman and Melinda A. Roberts, 289–314. Springer, 2009.

———. *Population Ethics: The Challenge of Future Generations*, unpublished manuscript.

Arrhenius, Gustaf, and Wlodek Rabinowicz. “The Value of Existence.” *The Oxford Handbook of Value Theory*, 2015, 424–443.

Ashford, Elizabeth. “The Demandingness of Scanlon’s Contractualism.” *Ethics* 113, no. 2 (2003): 273–302.

Bader, Ralf M. *Person-Affecting Population Ethics*, unpublished manuscript.

———. “Person-Affecting Utilitarianism.” In *Oxford Handbook of Population Ethics*, forthcoming.

Boonin, David. *The Non-Identity Problem and the Ethics of Future People*. Oxford University Press, 2014.

Brock, D. W., and D. Wikler. “Ethical Challenges In Long-Term Funding For HIV/AIDS.” *Health Affairs* 28, no. 6 (November 1, 2009): 1666–76.

Broome, John. “Equality versus Priority: A Useful Distinction.” *Economics and Philosophy* 31, no. 02 (July 2015): 219–28.

———. *Ethics out of Economics*. Cambridge: Cambridge University Press, 1999.

———. “Fairness.” In *Proceedings of the Aristotelian Society*, 91:87–101, 1990.

———. “General and Personal Good.” In *The Oxford Handbook of Value Theory*, edited by Iwao Hirose and Jonas Olson, 249–66. Oxford University Press, 2015.

———. “Kamm on Fairness.” *Philosophy and Phenomenological Research* 58, no. 4 (1998): 955–61.

———. “Normative Requirements.” *Ratio* 12, no. 4 (December 1999): 398–419.

———. “Reasons.” In *Reason and Value : Themes from the Moral Philosophy of Joseph Raz*, edited by Philip Pettit, R Jay Wallace, Sam Scheffler, and Michael Smith. New York: Oxford University Press, 2004.

———. “The Public and Private Morality of Climate Change.” *The Tanner Lectures on Human Values*, 2012.

———. “Uncertainty and Fairness.” *The Economic Journal* 94, no. 375 (September 1984): 624.

———. *Weighing Goods*. Cambridge, Mass.: Basil Blackwell, 1991.

———. *Weighing Lives*. Oxford University Press, 2004.

Bykvist, Krister. “Being and Wellbeing.” In *Weighing and Reasoning: Themes from the Philosophy of John Broome*, edited by Iwao Hirose and Andrew Evan Reisner, First edition., 87–94. Oxford ; New York, NY: Oxford University Press, 2015.

———. “The Benefits of Coming into Existence.” *Philosophical Studies* 135, no. 3 (September 8, 2007): 335–62.

Carlson, Erik. *Consequentialism Reconsidered*. Kluwer Academic Publishers, 1995.

Casal, Paula. “Why Sufficiency Is Not Enough.” *Ethics* 117, no. 2 (2007): 296–326.

Crisp, Roger. “Equality, Priority, and Compassion.” *Ethics* 113, no. 4 (2003): 745–763.

Diamond, Peter A. “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment.” *Journal of Political Economy* 75, no. 5 (1967): 765–66.

Dorsey, Dale. “Headaches, Lives and Value.” *Utilitas* 21, no. 01 (March 2009): 36–58.

Dougherty, Thomas. “Aggregation, Benificence and Chance.” *Journal of Ethics & Social Philosophy. Vol. 7, No. 2*, 2013.

Elster, John. *Solomonic Judgments: Studies in the Limitations of Rationality*. Cambridge University Press, 1989.

Epstein, Larry G., and Uzi Segal. “Quadratic Social Welfare Functions.” *Journal of Political Economy* 100, no. 4 (August 1992): 691–712.

Ewing, A. C. *The Definition of Good*. London: Routledge and Kegan Paul, 1947.

Fleurbaey, Marc. “Two Variants of Harsanyi’s Aggregation Theorem.” *Economics Letters* 105, no. 3 (December 2009): 300–302.

———. “Welfare Economics, Risk and Uncertainty.” *Canadian Journal of Economics/Revue Canadienne d’économique* 51, no. 1 (February 2018): 5–40.

Fleurbaey, Marc, and Alex Voorhoeve. “Decide As You Would with Full Information!” In *Inequalities in Health*, edited by Nir Eyal, Samia A. Hurst, Ole F. Norheim, and Dan Wikler, 113–28. Oxford University Press, 2013.

———. “On the Social and Personal Value of Existence.” In *Weighing and Reasoning: Themes from the Philosophy of John Broome*, edited by Iwao Hirose and Andrew Reisner, 95–109. Oxford University Press, 2015.

Frankfurt, Harry. "Equality as a Moral Ideal." *Ethics* 98, no. 1 (1987): 21–43.

Frick, Johann. "Context-Dependent Betterness and the Mere Addition Paradox," unpublished manuscript.

———. "Contractualism and Social Risk." *Philosophy & Public Affairs* 43, no. 3 (June 2015): 175–223.

———. "On the Survival of Humanity." *Canadian Journal of Philosophy* 47, no. 2–3 (May 4, 2017): 344–67.

———. "Treatment versus Prevention in the Fight against HIV/AIDS and the Problem of Identified versus Statistical Lives." In *Identified versus Statistical Lives: An Interdisciplinary Perspective*, edited by I. Glenn Cohen, Norman Daniels, and Nir Eyal. Oxford University Press, 2015.

———. "Uncertainty and Justifiability to Each Person." In *Inequalities in Health*, edited by Nir Eyal, Samia A. Hurst, Ole F. Norheim, and Dan Wikler, 129–46. Oxford University Press, 2013.

Fried, Barbara H. "Can Contractualism Save Us from Aggregation?" *The Journal of Ethics* 16, no. 1 (2012): 39–66.

Greaves, Hilary. "Antiprioritarianism." *Utilitas* 27, no. 01 (March 2015): 1–42.

———. "Population Axiology." *Philosophy Compass* 12, no. 11 (November 2017): 1–15.

Greaves, Hilary, and William MacAskill. "The Case for Strong Longtermism." *GPI Working Paper 7*, 2019.

Halstead, John. "The Numbers Always Count." *Ethics* 126, no. 3 (April 2016): 789–802.

Hare, Caspar. "Should We Wish Well to All?" *Philosophical Review* 125, no. 4 (October 2016): 451–72.

Harsanyi, John C. “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility.” *Journal of Political Economy* 63, no. 4 (1955): 309–21.

Henning, Tim. “From Choice to Chance? Saving People, Fairness, and Lotteries.” *Philosophical Review* 124, no. 2 (April 2015): 169–206.

Holtug, Nils. “On the Value of Coming into Existence.” *The Journal of Ethics* 5, no. 4 (2001): 361–384.

Horton, Joe. “Aggregation, Complaints, and Risk.” *Philosophy & Public Affairs* 45, no. 1 (2017): 54–81.

———. “Aggregation, Risk, and Reductio.” *Ethics*, forthcoming.

———. “Always Aggregate.” *Philosophy & Public Affairs* 46, no. 2 (March 2018): 160–74.

Kagan, Shelly. *The Limits of Morality*. Oxford University Press, 1989.

Kamm, F. M. *Intricate Ethics*. Oxford University Press, 2007.

———. *Morality, Mortality Volume I: Death and Whom to Save From It*. Oxford University Press, 1998.

———. “Nonconsequentialism.” In *The Blackwell Guide to Ethical Theory*, edited by Hugh LaFollette and Ingmar Persson, 2nd ed. Blackwell Publishing, 2013.

Kornhauser, Lewis. A., and Lawrence G. Sager. “Just Lotteries.” *Social Science Information* 27, no. 4 (1988): 483–516.

Kowalzcyk, Kacper. “Yet Another Argument Against Anti-Aggregation,” unpublished manuscript.

Kumar, Rahul. “Risking and Wronging.” *Philosophy & Public Affairs* 43, no. 1 (2015): 27–51.

Lewis, C.S. *The Problem of Pain*. Collins/Fontana Books, 1957.

Lewis, David. “A Subjectivist’s Guide to Objective Chance.” In *Studies in Inductive Logic And*, edited by Richard Jeffrey, 267–297. University of America Press, 1980.

Loewenstein, George, and Karen Jenni. “Explaining the ‘Identifiable Victim Effect.’” *Journal of Risk and Uncertainty* 14 (1997): 235–257.

Marschak, Jacob. “Rational Behavior, Uncertain Prospects, and Measurable Utility.” *Econometrica* 18, no. 2 (1950): 11–141.

McClennen, Edward F. *Rationality and Dynamic Choice: Foundational Explorations*. 1st ed. Cambridge University Press, 1990.

McKerlie, Dennis. “Egalitarianism.” *Dialogue* 23 (1984): 223–37.

McMahan, Jeff. “Asymmetries in the Morality of Causing People to Exist.” In *Harming Future Persons*, edited by Melinda A. Roberts and David T. Wasserman, 35:49–68. Dordrecht: Springer Netherlands, 2009.

———. “Causing People to Exist and Saving People’s Lives.” *The Journal of Ethics* 17, no. 1–2 (June 2013): 5–35.

Mogensen, Andreas L. “Staking Our Future: Deontic Longtermism and the Non Identity Problem.” *GPI Working Paper 9*, 2019, 32.

Moore, R. “Caring for Identified versus Statistical Lives: An Evolutionary View of Medical Distributive Justice.” *Ethology and Sociobiology* 17, no. 6 (1996): 379–401.

Moulin, Hervé. “Condorcet’s Principle Implies the No Show Paradox.” *Journal of Economic Theory* 45, no. 1 (June 1988): 53–64.

Munoz-Dardé, Véronique. “The Distribution of Numbers and the Comprehensiveness of Reasons.” *Proceedings of the Aristotelian Society* 105 (2005): 191–217.

Myerson, Roger B. “Utilitarianism, Egalitarianism, and the Timing Effect in Social Choice Problems.” *Econometrica* 49, no. 4 (July 1981): 883.

Nagel, Thomas. “Equality.” In *Mortal Questions*, 106–28. Cambridge: Cambridge University Press, 1979.

Narveson, Jan. “Utilitarianism and New Generations.” *Mind* 76 (1967): 62–72.

Norcross, Alastair. "Comparing Harms: Headaches and Human Lives." *Philosophy & Public Affairs* 26, no. 2 (1997): 135–67.

———. "Great Harms from Small Benefits Grow: How Death Can Be Outweighed by Headaches." *Analysis* 58, no. 2 (1998): 152–158.

———. "Intransitivity and the Person-Affecting Principle." *Philosophy and Phenomenological Research* 59, no. 3 (1999): 769–76.

Ord, Toby. "A New Counterexample to Prioritarianism." *Utilitas* 27, no. 03 (September 2015): 298–302.

Otsuka, Michael. "How It Makes a Moral Difference That One Is Worse off than One Could Have Been." *Politics, Philosophy & Economics* 17, no. 2 (May 2018): 192–215.

———. "Prioritarianism and the Measure of Utility." *Journal of Political Philosophy* 23, no. 1 (March 2015): 1–22.

———. "Risking Life and Limb: How to Discount Harms by Their Improbability." In *Identified versus Statistical Lives: An Interdisciplinary Perspective*, edited by I. Glenn Cohen, Norman Daniels, and Nir Eyal. Oxford University Press, 2015.

———. "Saving Lives, Moral Theory, and the Claims of Individuals." *Philosophy & Public Affairs* 34, no. 2 (2006): 109–135.

———. "Scanlon and the Claims of the Many versus the One." *Analysis* 60, no. 3 (2000): 288–93.

Otsuka, Michael, and Alex Voorhoeve. "Equality Versus Priority." In *The Oxford Handbook of Distributive Justice*, edited by Serena Olsaretti, Vol. 1. Oxford University Press, 2018.

———. "Why It Matters That Some Are Worse off than Others: An Argument against the Priority View." *Philosophy & Public Affairs* 37, no. 2 (2009): 171–199.

Parfit, Derek. "Equality and Priority." *Ratio (New Series)* X, no. 3 (1997): 202–21.

———. "Future People, the Non-Identity Problem, and Person-Affecting Principles." *Philosophy & Public Affairs* 45, no. 2 (March 2017): 118–57.

———. "Justifiability to Each Person." *Ratio (New Series)* XVI, no. 4 (2003): 368–90.

———. *Reasons and Persons*. Oxford University Press, 1984.

Persson, Ingmar. "Equality, Priority and Person-Affecting Value." *Ethical Theory and Moral Practice* 4, no. 1 (2001): 23–39.

Pummer, Theron. "Whether and Where to Give." *Philosophy & Public Affairs* 44, no. 1 (December 2016): 77–95.

Rawls, John. *A Theory of Justice (Revised Edition)*. Oxford University Press, 1999.

Regan, Donald H. "Why Am I My Brother's Keeper?" In *Reason and Value – Themes from the Moral Philosophy of Joseph Raz*, edited by R Jay Wallace, Philip Pettit, Sam Scheffler, and Michael Smith, 202–30. Oxford: Oxford University Press, 2004.

Reibetanz, Sophia. "Contractualism and Aggregation." *Ethics* 108, no. 2 (January 1998): 296–311.

Ridge, Michael. "How to Avoid Being Driven to Consequentialism: A Comment on Norcross." *Philosophy & Public Affairs* 27, no. 1 (1998): 50–58.

Ross, Jacob. "Rethinking the Person-Affecting Principle." *Journal of Moral Philosophy* 12, no. 4 (August 11, 2015): 428–61.

Rüger, Korbinian. "Aggregation with Constraints." *Utilitas*, forthcoming.

———. "On Ex Ante Contractualism." *Journal of Ethics and Social Philosophy* 13, no. 3 (2018).

Savage, Leonhard. *The Foundations of Statistics*. 2nd ed. Dover, 1972.

Scanlon, T.M. "Contractualism and Utilitarianism." In *Utilitarianism and Beyond*, edited by Amartya Sen and Bernard Williams, 103–29. Cambridge University Press, 1982.

———. "Reply to Zofia Stemplowska." *Journal of Moral Philosophy* 10 (2013): 508–514.

———. *What We Owe to Each Other*. Cambridge, Mass. ; London: Belknap Press of Harvard University Press, 1998.

Schelling, T. "The Life You Save May Be Your Own." In *Problems in Public Expenditure Analysis*, edited by S.B. Chase, 127–62. Brookings Institution Washington D.C., 1968.

Sen, Amartya. "Internal Consistency of Choice." *Econometrica* 61, no. 3 (1993): 495–521.

Sinclair, Thomas. "Are We Conditionally Obligated to Be Effective Altruists?" *Philosophy & Public Affairs* 46, no. 1 (January 2018): 36–59.

Starmer, Chris. "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk." *Journal of Economic Literature* 38, no. 2 (June 2000): 332–82.

Steuwer, Bastian. "Contractualism, Complaints, and Risk," unpublished manuscript.

Stocker, Michael. "Agent and Other: Against Ethical Universalism." *Australasian Journal of Philosophy* 54, no. 3 (December 1976): 206–20.

Tadros, Victor. "Controlling Risk." In *Prevention and the Limits of the Criminal Law*, edited by Patrick Tomlin, Andrew Ashworth, and Lucia Zedner. Oxford University Press, 2013.

———. "Localized Restricted Aggregation." *Oxford Studies in Political Philosophy* 5 (2019): 171–204.

Taurek, John M. "Should the Numbers Count?" *Philosophy & Public Affairs* 6, no. 4 (1977): 293–316.

Temkin, Larry. "Equality as Comparative Fairness." *Journal of Applied Philosophy* 34, no. 1 (February 2017): 43–60.

Temkin, Larry S. "Equality, Priority or What?" *Economics and Philosophy* 19, no. 1 (April 2003): 61–87.

———. *Inequality*. Oxford University Press, 1993.

———. "Intransitivity and the Mere Addition Paradox." *Philosophy & Public Affairs*, 1987, 138–187.

———. *Rethinking the Good : Moral Ideals and the Nature of Practical Reasoning*. New York: Oxford University Press, 2015.

Tomlin, Patrick. "On Limited Aggregation." *Philosophy & Public Affairs* 45, no. 3 (June 2017): 232–60.

Tungodden, Bertil, and Peter Vallentyne. "On the Possibility of Paretian Egalitarianism." *The Journal of Philosophy* 102, no. 3 (2005): 126–154.

Voorhoeve, Alex. "How Should We Aggregate Competing Claims?" *Ethics* 125, no. 1 (October 2014): 64–87.

———. "Why One Should Count Only Claims with Which One Can Sympathize." *Public Health Ethics*, March 3, 2016, 1–9.

Voorhoeve, Alex, and Marc Fleurbaey. "Egalitarianism and the Separateness of Persons." *Utilitas* 24, no. 03 (September 2012): 381–98.

Walen, Alec. "Risks and Weak Aggregation - Why Different Models of Risk Suit Different Types Cases." *Ethics*, forthcoming.

Wasserman, David. "Let Them Eat Chances: Probability and Distributive Justice." *Economics and Philosophy* 12, no. 01 (1996): 29–49.