

London School of Economics and Political Science

One-by-One:
Moral Theory for Separate Persons

Bastian Steuwer

A thesis submitted to the Department of Philosophy, Logic and
Scientific Method of the London School of Economics and
Political Science for the degree of Doctor of Philosophy,
London, 14th May 2020.

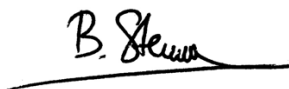
Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others.

I confirm that a version of Chapter 1 is published as “Why It Does Not Matter What Matters: Personal Identity, Relation R, and Moral Theory” in *The Philosophical Quarterly* 70 (2020): 178-98; a version of Chapter 3 is forthcoming as “Contractualism, Complaints, and Risk” in *Journal of Ethics and Social Philosophy*; and a version of Chapter 5 is forthcoming as “Aggregation, Balancing, and Respect for the Claims of Individuals” in *Utilitas*.

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that this thesis consists of 84,347 words.

A handwritten signature in black ink, appearing to read "B. Steuwer", is written above a horizontal line.

Bastian Steuwer

Abstract

You and I lead different lives. While we share a society and a world, our existence is separate from one another. You and I matter individually, by ourselves. My dissertation is about this simple thought. I argue that this simple insight, the separateness of persons, tells us something fundamental about morality. My dissertation seeks to answer how the separateness of persons matters. I develop a precise view of the demands of the separateness of persons. The separateness of persons imposes both a requirement on the justification of first-order moral principles as well as a requirement on the content of first-order moral principles. In specifying these demands, I argue that respecting the separateness of persons requires taking into consideration each person's point of view separately. This requires taking into account the moral relations in which individuals stand to one another. I make use of this relational understanding of the separateness of persons to advance various debates in moral and political philosophy. I argue for a framework to assess to which extent the veil of ignorance can be reconciled with the separateness of persons. I also argue for a new view on the ethics of risk which is a form of contractualism that discounts risks only by their objective risk. Furthermore, I argue for a new solution to the problem of aggregation that is skeptical of aggregation and can set plausible limits to aggregation. Lastly, I provide a new relational agent-based justification for deontological constraints. In addition to answering how the separateness of persons matters, I defend the separateness of persons against challenges. Most importantly, I argue that the importance of the separateness of persons is not undermined even if we believe that our personal identity, i.e. whether we persist as the same person, is unimportant.

Acknowledgments

My greatest thanks belong to my supervisors Michael Otsuka and Alex Voorhoeve for their invaluable support. I am thankful not only for countless discussions and feedback, but also for their encouragement throughout my PhD. Mike and Alex always had an open ear whenever I needed advice or support, philosophical or otherwise. Their feedback was always thorough, critical, and supportive. The process of discussing my ideas over and over with them has made me a better philosopher; not only better than I was at the beginning of my PhD but also better than I could have become had I studied at any other place. I learned a great amount from working with them. I can only hope that through my dissertation which all too often engages critically with their ideas and arguments I was able to partly return this favor.

The Philosophy Department as a whole has been a great place during these four years. I have felt at home in the Department and will greatly miss it. My thanks to the administrative support team that has been very helpful and supportive throughout the time, and especially to Andrea Pawley for her tireless help with the (ultimately successful) job market. Faculty members of the Department have been extraordinarily generous with their time for a student whom they did not supervise. Richard Bradley, Christian List, and Anna Mahtani have been outstanding in helping me better understand areas of philosophy outside my own specialty. Susanne Burri, Peter Dennis, and Jonathan Parry took time to read my work and give feedback. Jonathan Birch has been very supportive with all job market related affairs.

I also have been blessed with a terrific crowd of PhD students whose solidarity and friendship made my time in the PhD more enjoyable and special. An extra thanks belongs to those in my cohort: Chloé de Canson, David Coombs, Paul Daniell, Christina Easton, Ko-Hung Kuan, and Joe Roussos; and to those outside my cohort who took the time to read my work and whose comments greatly improved the final dissertation: Nicolas Côté, Todd Karhu, Chris Marshall, and Tom Rowe.

Over the course of my PhD, I presented parts of my dissertation at various venues. I am thankful to the audience at Fribourg, Sheffield, UCL, Barcelona, KCL, Oxford,

Düsseldorf, Stockholm, Salzburg, and the LSE Choice Group (twice), including my commentators at Stockholm, Nicolas Olsson-Yaouzis, and Oxford, Roger Crisp. Thanks also to those people outside of LSE who took time to read parts of my dissertation and gave comments on it: Tomi Francis, Johann Frick, Joe Horton, Thulasi K. Raj, Korbinian Rieger, Joachim Wündisch, as well as anonymous journal referees for (too many) journals.

For their constant love and support I thank my German and Indian families as well as my friends scattered across the world in Seattle, Ann Arbor, Tacámbaro, Barcelona, London, Paris, Geneva, The Hague, Brussels, Berlin, Jeddah, Delhi, Bombay, Kochi, Madurai, Shanghai and places I have forgotten. Without their support and love I could not have achieved this project.

My last acknowledgment is to the person whom I owe the most. When I started my PhD, I had to say goodbye to you, Thulasi, not knowing when I would see you again. My greatest project over the last four years was not to complete my PhD, it was to overcome all of our obstacles and build a life together. I can now say with great pride that we succeeded against all odds and adversity. I cannot express my gratitude for all your love and support, and I cannot express my joy in looking forward to what comes next. This dissertation is dedicated to you.

Table of Contents

Introduction	10
I. Different Trade-Offs	11
II. First-Person Standpoint: Anti-Aggregation	13
III. Second-Person Standpoint:	
The Separateness and Relatedness of Persons	18
IV. Two Versions of the Separateness of Persons Objection	24
V. The Justificatory Requirement	27
VI. The Substantive Requirement	29
VII. The Separateness of Persons, Attitudes, and Rightness	33
VIII. The Dissertation	36
Chapter 1. Why It Does Not Matter What Matters:	
Relation R, Personal Identity, and Moral Theory	41
I. Introduction	41
II. First Argument: Less United Individuals	43
III. Second Argument: Less Separate Persons	54
IV. Third Argument: Less Importance to Persons	60
V. Conclusion	67
Chapter 2. Separate Persons Behind the Veil	69
I. Introduction	69
II. Justification for Separate Persons	70
A. The Impartial Spectator	70
B. John Harsanyi	71
C. John Rawls	78
D. From Rawls to Dworkin: The Nature of the Veil	81
E. From Harsanyi and Rawls to Dworkin:	
The Justificatory Role of the Veil	82
F. From Rawls to Dworkin: Collective Assets	89
G. Summary	93

III.	Principles for Separate Persons	94
A.	Harsanyi's Average Utilitarianism	94
B.	Rawls's Two Principles of Justice	96
C.	Dworkin's Equality of Resources	97
IV.	Conclusion	101
Chapter 3.	Contractualism, Complaints, and Risk	103
I.	Contractualism and Risk	103
II.	Otsuka's Sequence	105
III.	A Problem for Ex Post Contractualism	114
IV.	What We Owe ... to Whom?	119
A.	Justifiability to Each Separate Person	120
B.	The Luckless and the Doomed	126
V.	Objections	133
A.	Determinism	133
B.	Ex Ante Pareto	138
C.	Identified Victim Bias	140
VI.	Conclusion	144
Chapter 4.	Skepticism about Aggregation and Uncertain Rescues	146
I.	Ex Post Claims	148
II.	First Option: Relevance Tied to a State of the World	150
III.	Second Option: Relevant Inside and Across States of the World	153
IV.	Ex Post Limited Aggregation Without Ex Post Claims	157
Chapter 5.	Aggregation, Balancing, and Respect for the Claims of Individuals	159
I.	Introduction	159
II.	Relevance and Limited Aggregation	161
III.	Justifying Hybrid Balance Relevant Claims	164
IV.	Illustrations of Hybrid Balance Relevant Claims	175

V.	Hybrid Balance Relevant Claims and Objections to Limited Aggregation	179
	A. Problems with Global Relevance	180
	B. Problems with Local Relevance	183
	C. Principle of Agglomeration	186
VI.	Conclusion	189
	Balancing Three (or More) Groups	191
	Hybrid Balance Relevant Claims versus Sequential Matching	193
	The Order of Balancing	195
	Chapter 6. Constraints, You, and Your Victims	198
I.	Introduction	198
II.	A Relational Agent-Based Justification for Side Constraints	201
III.	The Puzzle of Minimizing One's Own Violations	207
	A. One's Own Violations and the <i>Guilty Agent</i>	207
	B. Responding to the <i>Guilty Agent</i>	209
	C. First Step: Choice between Doing and Allowing	210
	D. Second Step: Killing versus Letting Be Killed	213
	E. Second Step: Killing versus Letting Be Killed by Oneself	214
IV.	Why Evaluate Actions One at a Time?	219
V.	Self-Indulgence	223
VI.	Constraints and Non-Persons	224
VII.	The Next Constraint You Come Up Against	226
	Bibliography	228

List of Tables

Table 4.1 Case One	176
Table 4.2 Case Two	176
Table 4.3 Case Three	177
Table 4.4 Variation on Case Two	178
Table 4.5 Anchor by Competition	180
Table 4.6 Anchor by Strength	181
Table 4.7 Horton against Local Relevance	185
Table 4.8 Principle of Agglomeration	187
Table 4.9 Principle of Agglomeration (Alternative)	187
Table 4.10 Balancing Three Groups	192
Table 4.11 Hybrid Balance Relevant Claims versus Sequential Matching	194
Table 4.12 The Order of Balancing	195
Table 4.13 Hybrid Balance Relevant Claims versus Strongest Decides	196
Table 4.14 Hybrid Balance Relevant Claims versus Strongest Decides (II)	196

Introduction

Suppose you have one child that suffers from a painful disability. There is a treatment available that will remove the disability, but the treatment is arduous for your child. The treatment requires you to move to the city. Your child will lose their current friends and have no longer the joy of nature that provided relief from the disability. But after the treatment is over your child will flourish more without the disability. Are you permitted to move to the city to gain access to the treatment? The answer it seems depends on whether the long-term benefits outweigh the short-term losses. If on balance they do, it is permissible for you to proceed. If on balance they do not, it is impermissible for you to proceed. Now suppose you have two children, one of which suffers from a painful disability. The treatment requires you to move to the city. But now the burdens are on your able-bodied child instead of your disabled child. Are you permitted to move to the city to access the treatment? Here it seems that the answer is more difficult. Plausibly, you are allowed to move to the city even if on balance the burdens to your able-bodied child somewhat outweigh the benefits to your disabled child.¹ Why is this? Why is the answer in the first case not available in the second case?

The reason is that the second case is a trade-off between the interests of two persons whereas the trade-off in the first case is a trade-off between different interests of one person. The second trade-off is *inter-personal*, the first trade-off is *intra-personal*. But why can we not simply proceed the same way for inter-personal trade-offs as we do for intra-personal trade-offs? To treat both trade-offs the same way would violate the *separateness of persons*.

You and I lead different lives. While we share a society and a world, our existences are separate from one another. You and I matter individually, by ourselves. This is the core idea, the simple thought, behind the separateness of persons. In this

¹ The contrast is inspired by a case of Thomas Nagel's. Thomas Nagel, *Mortal Questions* (Cambridge: Cambridge University Press, 1979), pp. 123-24. A similar one-child case, albeit under conditions of risk, is presented and contrasted with Nagel's two-child case by 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 (2009): 171-99, at p. 188.

dissertation, I argue that this simple insight tells us something fundamental about morality.

I. Different Trade-Offs

The idea that different kinds of trade-offs require different solutions is not a new one. The earliest version of this criticism that I have been able to locate is due to Richard Price. It occurs in Price's book *A Review of the Principal Questions in Morals* in 1758. Price examines the view that the sole standard for justice is general utility or public happiness. In effect, the view that he examines is a utilitarian theory of justice. Price does not attribute this view to any specific author and his criticism predates the beginning of classical utilitarianism with Bentham's *An Introduction to the Principles of Morals and Legislation* in 1789. In a footnote, Price mentions, however, an affinity between the view he examines and the work of Frances Hutcheson.²

Price begins by pointing out how under the view in question people may be put into misery if this is what overall happiness demands. He then asks that from the standpoint of such a utilitarian principle:

“What makes the difference between communicating happiness to a *single being* in such a manner, as that it shall be only the excess of his enjoyments above his sufferings; and communicating happiness to a *system of beings* in such a manner that a *great* number of them shall be totally miserable, but a *greater* number happy? Would there be nothing in such a procedure that was not right and just ... Such consequences are plainly shocking to our natural sentiments; but I know not how to avoid them on the principles I am examining.”³

² Richard Price, *A Review of the Principal Questions in Morals* (Oxford: Clarendon Press, 1948), p. 161fn. Hutcheson is a precursor to classical utilitarianism whose moral thought shows great resemblance to utilitarianism. He both gives an analysis of rights in terms of their tendency to promote the universal good and provides a criterion for evaluating actions that resembles Bentham's greatest happiness principle. See Julia Driver, “The History of Utilitarianism,” in *The Stanford Encyclopedia of Philosophy*. Winter 2014 Edition, ed. Edward N. Zalta (URL: <<https://plato.stanford.edu/archives/win2014/entries/utilitarianism-history/>>)

³ Price, *A Review of the Principal Questions in Morals*, p. 160. Emphasis in the original.

Price observes that a utilitarian principle must ignore the difference between balancing the happiness and suffering of a single being and balancing the happiness and suffering of a system of beings. In effect, Price describes here that a utilitarian principle can draw no distinction between inter-personal and intra-personal trade-offs. To treat these two trade-offs alike is shocking to Price and an objection to any principle that entails it.

Price approaches the distinction between different kinds of trade-offs by way of a criticism of a view that fails to distinguish between them. R.B. Perry, who is another early philosopher that insisted on difference between inter-personal and intra-personal trade-offs, approached the subject from a different angle. In a section of his book *General Theory of Value* (1926) titled “The Independence of Persons” Perry approaches the difference between trade-offs as a problem of integrating different interests.⁴ Perry argues that there is a difference between personal integration (resolving intra-personal trade-offs) and social integration (resolving inter-personal trade-offs). He conceives of interests as representing distinct ends. In the case of resolving conflicts of ambivalence between our own interests, we can resolve them by subsuming our different interests under one end. We should treat our various individual interests as means to one overarching end. This he called the “principle of subordination”. This principle, however, is inapplicable in the case of social integration. If I, as a decision-maker, treat someone else’s interests only as means to my own ends, then I treat this person as a mere means. Different persons, however, have their separate ends. Perry attributes the mistake of overlooking this difference to a tendency to personify society and treat it as a singular subject. Unlike Price, Perry does not, however, mention any specific moral theory as guilty of overlooking the difference between individual, intra-personal trade-offs and social, inter-personal trade-offs.

More recently, David Gauthier has raised the criticism that a theory that assimilates the two kinds of trade-offs overlooks the separateness of persons. Gauthier’s criticism makes explicit appeal to the idea of the separateness of persons

⁴ Ralph Barton Perry, *General Theory of Value* (Cambridge, MA.: Harvard University Press, 1926), pp. 674-77.

which was merely implicit in the arguments of Price and Perry. Gauthier's argument in his *Practical Reasoning* (1963) is therefore among the first clear invocations of the separateness of persons in moral argument.⁵ Gauthier remarks that whenever prudence is concerned, one is not interested in the satisfaction of one's desires by themselves, but rather in one's own greater satisfaction of one's desires. The separateness of these desires does not matter. Things are different, however, in the case where conflicting desires of different persons come into play. Here we need to pay special attention to the individual desires and not only to the sum-total of all desires. Doing otherwise, Gauthier proceeds to argue, would mean that one considers the different desires of different persons to be part of one system of desires. But no such super-person exists. It is individuals that have desires.

II. First-Person Standpoint: Anti-Aggregation

But why exactly should we treat the two kinds of trade-offs differently? Gauthier seems to suggest that to do otherwise would be to treat humanity as one super-person. Yet it is patently obvious to adherents of utilitarianism that humanity is not one super-person. The arguments for utilitarianism do not typically assume humanity to be a super-person.⁶ Perry's reason was that different persons are different ends in themselves. We can accept Perry's reason while still maintaining that we need some theory that tells us how to trade-off goods that accrue to different final ends. What rules out that this theory is the same as the theory for intra-personal trade-offs?⁷ Price rejected the equivalence of the two kinds of trade-offs because it implies unacceptable conclusions. But counterintuitive implications of utilitarianism are already well-known. The inability of distinguishing between different trade-offs was supposed to be an additional and theoretical argument.

⁵ David Gauthier, *Practical Reasoning* (Oxford: Clarendon Press, 1963), pp. 123-27. As we shall see later, John Rawls develops his version of the separateness of persons at the same time as Gauthier.

⁶ See also Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), p. 331.

⁷ This is in effect the point of Richard Yetter Chappell's utilitarian response to the separateness of persons objection. Richard Yetter Chappell, "Value Receptacles," *Noûs* 49 (2015): 322-32.

One answer to the question why the difference between the trade-offs matters lies in a link we can draw to a line of opposition to aggregative reasoning. What seems to make the initial decision of moving to the city or staying in the suburbs difficult in the two children case is that it seems wrong to simply aggregate all benefits and burdens in this case. This link between the separateness of persons and opposition to aggregation also has an important historical pedigree.

A discussion by the lay theologian C.S. Lewis in *The Problem of Pain* (1940) exemplifies nicely this thought.⁸ Lewis's argument, however, has also given grounds for a reaction skeptical of the importance of the separateness of persons. As a theologian Lewis was interested in responding to the problem of evil. He gives the following argument to establish that the existence of evil and suffering is less widespread than one might think. Imagine a person who has a toothache of a given intensity. Now imagine a second person with a toothache of the same intensity. A natural thought would be that the pain in the world has doubled. But Lewis resists this thought. There is no one who suffers a toothache twice as intense. Suffering does not add up in this way. From this he concludes that "[when] we have reached the maximum that a single person can suffer, we have ... reached all the suffering there ever can be in the universe".⁹ Lewis's thought is that we cannot simply aggregate harms in a simple manner because there is no agent who will be the subject of this harm. This indicates a clear difference to cases of intra-personal aggregation where all harms and benefit fall on one person. While Lewis gives a reason for distinguishing between intra-personal and inter-personal aggregation, the argument has attracted opposition.¹⁰ Lewis's further claim that all the suffering in the universe is exhausted by the worst that a single person can suffer overreaches. It takes the thought that only individuals matter to an extreme. This invites the suspicion that the separateness of persons leads to implausible moral demands which are entirely focused on the fates of single individuals.

⁸ C.S. Lewis, *The Problem of Pain* (Québec: Samizdat University Press, 2016), pp. 72-73.

⁹ Lewis, *The Problem of Pain*, p. 73.

¹⁰ See e.g. Derek Parfit, "Innumerate Ethics," *Philosophy & Public Affairs* 7 (1978): 285-301, at pp. 294-96.

A more moderate version of the idea that aggregating individual goods differs in the contexts of different trade-offs comes from John Findlay in his book *Values and Intentions* (1961).¹¹ Findlay notes an asymmetry between the aggregation of satisfactions within a person's life and across different person's lives. In particular, Findlay rejects a simple aggregative model for satisfactions within one's life. He draws a distinction between totals of satisfaction and a total satisfaction. These two can diverge. A holistic experience of an extended period of time is different from an additive total of the satisfactions at each point in time. A vacation with a bad ending may be spoiled because of it, while a vacation with a bad start may be remembered as a nice experience. These holistic experiences of total satisfaction should count in thinking about the value of our life. Such a model is, however, plainly not available for aggregating preferences across people's lives. There are only totals of satisfactions and no holistic experience of total satisfaction. There is no individual point of view from which this whole experience is made. This is due to "the profound gulfs constitutive of the space of persons".¹² Because of the separateness of different lives, there is no overall experience that we can give. This does not exclude the possibility of an overall assessment of the satisfactions of different people. Rather, the idea is that combining various benefits and burdens must proceed very differently in the case of individuals than in the case of collectives.

The more radical opposition to aggregation has become identified with the separateness of persons. The two modern references are Robert Nozick and John Taurek. In his *Should the Numbers Count?* (1977) Taurek argues against the view that we have a duty to save a greater number of people from equal harm rather than a lesser number.¹³ Taurek's skeptical argument against aggregation resembles Lewis's at some stages. For example, Taurek invokes the idea that it is a mistake to think that small pains experienced by many people could be as bad as pains of greater intensity or duration suffered by a single person.¹⁴ For Taurek this is because there is no perspective for whom the many small pains are worse than a pain of the single

¹¹ John Findlay, *Values and Intentions* (London: George Allen & Unwin, 1961), pp. 234-36.

¹² Findlay, *Values and Intentions*, p. 236.

¹³ John M. Taurek, "Should the Numbers Count?," *Philosophy & Public Affairs* 6 (1977): 293-316.

¹⁴ Taurek, "Should the Numbers Count?," pp. 307-10.

person. Taurek's claim about the badness of suffering is a claim about preferring one state of affairs to another. Saying that something is worse than something else is, for Taurek, tantamount to preferring that one state of affairs comes about rather than another.¹⁵ This is different from Lewis's claim that once a single person reaches maximum suffering, this exhausts all the suffering there possibly could be in the world. Lewis's claim is about the amount of suffering in the world. Taurek's argument focuses on what the rejection of aggregative reasoning means for the morality of saving from harm. Robert Nozick focuses in *Anarchy, State, and Utopia* (1974) on the question whether it can be permissible to impose harm on one person in order to bring about a greater social good. In a manner reminiscent of Lewis and Taurek, Nozick rejects that we can do so, because "there is no *social entity* with a good that undergoes some sacrifice for its own good". He continues:

"There are only individual people, different individual people, with their own individual lives. Using one of these people to benefit others, uses him and benefits others. Nothing more. ... To use a person in this way does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has. *He* does not get some overbalancing good from his sacrifice, and no one is entitled to force this upon him."¹⁶

For both Taurek and Nozick to aggregate overlooks the fact that different individuals have different points of view or first-personal standpoints. This distinguishes individuals from inanimate objects, for example. Taurek writes that to simply add up all benefits and burdens would mean we value persons the way we value objects.¹⁷ However, persons are not the only beings with first-personal standpoints. Conscious or sentient animals have a point of view. There are things that can be better or worse for them, they can be harmed, experience pain and pleasure.¹⁸

¹⁵ Taurek, "Should the Numbers Count?," pp. 304-5. See also Weyma Lübbe, "Taurek's No Worse Claim," *Philosophy & Public Affairs* 36 (2008): 68-85.

¹⁶ Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp. 32-33. Emphasis in the original.

¹⁷ Taurek, "Should the Numbers Count?," pp. 306-8.

¹⁸ Peter Godfrey Smith, for example, identifies "subjective experiences" or a "point of view" with the ability to feel experiences. If an animal can feel pain, then there is subjectivity and a

All of these are morally significant. But persons have capacities above and beyond this. Nozick writes that persons can have a conception of a good life. They can have plans of life, projects and strive towards a good life. They can have an idea of how they want to be, what sort of identity they want to adopt. Their lives can have meaning.¹⁹ This argument stresses that the separateness of persons says that individuals have separate lives *to lead*. It is individuals who are leading their lives and who have the opportunity to make something meaningful or valuable out of their lives. This opportunity means that we need to give greater importance to person's first-personal standpoints and cannot simply overlook them by aggregating across them.

Taurek's arguments seems to set strict limits to aggregation in the context of the morality of saving from harm. Nozick's argument seems to set strict limits to our ability to harm others. Critics of the separateness of persons have argued that the limits they set are too strict and that this gives us reason to doubt the importance of the separateness of persons. The complaint that individuals should not be sacrificed for the benefit of others is unreasonably general, according to these critics. Such an interpretation would result in an implausible Paretian morality which never requires anyone to even accept small sacrifices for the benefit of others. If morality can sometimes require us to balance the losses of some with the gains of others, then it appears that the appeal of the separateness of persons is mistaken here.²⁰ Proponents of the separateness of persons need a way to avoid such a Paretian morality without hollowing out the importance of the separateness of persons. A similar argument can be made about the rejection that the relative numbers matter in deciding whom to save. A morality that never allowed the relative numbers to count in deciding whom

point of view. Peter Godfrey-Smith, *Other Minds* (London: William Collins, 2017), ch. 4; and Peter Godfrey-Smith, "Evolving Across the Explanatory Gap," *Philosophy, Theory, and Practice in Biology* 11 (2019): 1-24.

¹⁹ Nozick, *Anarchy, State, and Utopia*, pp. 48-51.

²⁰ Parfit, *Reasons and Persons*, pp. 336-39; Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986), pp. 271-77; David O. Brink, "The Separateness of Persons, Distributive Norms, and Moral Theory," in *Value, Welfare, and Morality*, ed. R.G. Frey and Christopher Morris (Cambridge: Cambridge University Press, 1993), pp. 252-289, at pp. 253-59; and Larry Temkin, *Rethinking the Good* (Oxford: Oxford University Press, 2012), pp. 101-8; and Iwao Hirose, *Moral Aggregation* (Oxford: Oxford University Press, 2014), pp. 67-73.

to save seems overly restrictive according to these critics. Like the imagined Paretian morality, it fails to allow trade-offs that morality demands and gives an implausibly strong emphasis to single individuals.²¹ To counter this argument, proponents of the separateness of persons would need to show why the separateness of persons is not overly restrictive in this sense. In Chapters 3 (*Contractualism, Complaints, and Risk*), 4 (*Skepticism about Aggregation and Uncertain Rescues*) and 5 (*Aggregation, Balancing, and Respect for the Separateness of Persons*) I take up this challenge and develop a view that is guided by the separateness of persons while being neither Paretian nor holding that relative numbers are never morally relevant.

III. Second-Person Standpoint:

The Separateness and Relatedness of Persons

Thus far I argued that one of the reasons why overlooking the difference between inter-personal and intra-personal trade-offs is problematic is because this confuses what is permissible in aggregating benefits and burdens in one life with the permissibility of aggregation across lives. This is problematic because it conflates the different first-personal standpoints of individuals. Now, I want to suggest a second, equally important, reason why a moral theory needs to be sensitive to the difference between inter-personal and intra-personal trade-offs. This reason is related to the importance of person's second-personal authority and the importance of moral relations between persons.

The separateness of persons objection is often traced back to John Rawls's *A Theory of Justice* (1971). My previous discussion has already shown that the idea of the separateness of persons did not exclusively originate with Rawls's presentation of the objection. What is equally noteworthy is that the separateness of persons objection has antecedents in Rawls's own work. In *Justice as Fairness* (1958) Rawls compares

²¹ Brink, "The Separateness of Persons, Distributive Norms, and Moral Theory," pp. 259-82; and Alastair Norcross, "Two Dogmas of Deontology: Aggregation, Rights, and the Separateness of Persons," *Social Philosophy & Policy* 26 (2009): 76-95, at pp. 80-88.

utilitarianism to his own view of justice as fairness.²² His argument is not the familiar objection that utilitarianism can lead to counterintuitive verdicts or that it fails to account for the value of equal distributions. Rawls is willing to concede that what utilitarianism recommends might be extensionally equivalent to his own principles of justice. If all utility functions are identical, there is diminishing marginal utility, and costless redistribution, then utilitarianism would advocate for perfect equality of goods. Utilitarianism can, thereby, account for common sense principles of justice. Rawls rather objects that utilitarianism gives the wrong reason for accepting these principles of justice. The principles of justice would be accepted only as a response to the question of what the most efficient design of institutions is.²³

More interestingly Rawls objects that benefits to individuals matter only insofar as they contribute to the individual's welfare. Their importance is irrespective of any moral relations between individuals or any moral claims that they might be able to raise. Whether or not they are part of cooperative enterprises, for example, is immaterial. Rawls then criticizes the form of individualism that utilitarianism espouses. He writes:

“[Utilitarianism] regards persons as so many *separate* directions in which benefits and burdens may be assigned; and the value of the satisfaction and dissatisfaction of desire is not thought to depend in any way on the moral relations in which individuals stand, or on the kinds of claims which they are willing ... to press on each other.”²⁴

At first sight this criticism does not appear to be related to Rawls's criticisms that utilitarianism violates the separateness of persons. Rawls criticizes utilitarianism for admitting *too much* separation between individuals. This charge appears to be the precise opposite of the objection that utilitarianism overlooks the separation between individuals. Rawls's criticism in *Justice as Fairness* is that utilitarianism understands persons in an atomistic and unrelated manner. It fails to give importance to the

²² John Rawls, “Justice as Fairness,” *Philosophical Review* 67 (1958): 164-94, at pp. 184-87.

²³ G.A. Cohen has developed a somewhat similar criticism of Rawls's theory of justice as being concerned with “rules of regulation” for society as opposed to principles of justice in a fact-independent sense. See G.A. Cohen, *Rescuing Justice and Equality* (Cambridge, MA.: Harvard University Press, 2008), pt. 2.

²⁴ Rawls, “Justice as Fairness,” p. 187.

relations between individuals. In spite of the criticism that utilitarianism admits too much separation between persons, Rawls himself says that his separateness of persons objection takes its root from these considerations.²⁵ How is this possible? Considering the development of this idea in Rawls's thought helps us here.

Rawls develops his criticism that questions of justice are transformed into questions of efficient administration in *Constitutional Liberty and the Concept of Justice* (1963).²⁶ Rawls again compares utilitarianism (or social utility) with a view that takes justice as fundamental. The contrast here is that justice, in contrast to social utility, "takes the plurality of persons as fundamental". Social utility aims at maximizing one thing. Questions about social utility are therefore questions akin to efficient administration, namely questions of rational choice for a single chooser. Just as in the case of a single individual, losses to some part are immediately outweighed by gains to another. Justice forbids this kind of reasoning. The flaw of utilitarianism is to justify the violation of one person's claims by appeal to a compensating advantage that someone else has received. Rawls thereby singles out the importance of the competing claims of different individuals as one of the morally important relations which utilitarianism overlooks. To rectify the flaw of utilitarianism, Rawls proposes that we must find principles which can obtain the unanimous agreement of individuals. This agreement can be achieved from a position of equal liberty within moral constraints, i.e. Rawls's original position.

In *A Theory of Justice*, Rawls then gives the separateness of persons objection its famous form.²⁷ By extending the principle of rational choice for one person to social trade-offs, Rawls argues, utilitarianism fails to take seriously the separateness of persons. Again, Rawls picks up the criticism that utilitarianism overlooks the importance of moral relations. His proposal of "justice as fairness" is built around the recognition that it matters whether individuals are engaged in mutually advantageous cooperation or not. Rawls also gives another hint what the importance

²⁵ John Rawls, *A Theory of Justice*, rev. edn. (Oxford: Oxford University Press, 1999), p. 21fn10.

²⁶ John Rawls, "Constitutional Liberty and the Concept of Justice," in *John Rawls: Collected Papers*, ed. Samuel Freeman (Cambridge, MA.: Harvard University Press, 1999), pp. 73-95, at pp. 94-95.

²⁷ Rawls, *A Theory of Justice*, rev. edn., pp. 20-27.

of moral relations can mean. Moral relations matter, for example, in the way in which well-being arises. Well-being that is derived from discriminating against others should not be counted.²⁸ The importance of moral relations indicates that the moral importance of well-being is not irrespective of how it is created and how it bears on the relations between persons. In short, Rawls's argument here is a rejection of welfarism.

Rawls also gives a clearer answer to the question of why moral relations matter. The relations between persons matter insofar as they determine the appropriate principle of choice. Rawls writes that the right principle of regulation depends on what is regulated.²⁹ Principles of individual rationality are devised for single individuals. His principles of justice are devised for a plurality of individuals who all pursue their separate ends and who, moreover, are all part of a system of mutually advantageous cooperation. This makes clearer how Rawls's initial complaint that utilitarianism admits of too much separateness is connected to Rawls's later complaint that utilitarianism overlooks the separateness of persons. Principles of justice regulate the interactions of persons who are at the same time separate from and related to one another. If we use the same principle for inter-personal trade-offs that we use for intra-personal trade-offs, then we will overlook both aspects. We erode the distinction between persons and we also adopt an atomistic picture of human interaction. In its classical form, utilitarianism ignores the bonds between these atoms and justice is simply a function of the sum total of mass of these separate atoms. Utilitarianism, a theory which is insensitive to the difference between the different kinds of trade-offs, therefore, fails doubly.

The idea that overlooking the difference between different kinds of trade-offs primarily overlooks the importance of moral relations between individuals is also present in Thomas Nagel's discussion of the separateness of persons in *The Possibility of Altruism* (1970).³⁰ Nagel contrasts two different kinds of conflicts of reasons. One

²⁸ Rawls, *A Theory of Justice*, rev. edn., p. 27.

²⁹ Rawls, *A Theory of Justice*, rev. edn., p. 25.

³⁰ Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970), pp. 133-42. Nagel acknowledges that his thinking of the separateness of persons has taken root from Rawls's comments in *Constitutional Liberty and the Concept of Justice*. See Nagel, *The Possibility of Altruism*, p.134fn1.

conflict is between reasons which derive their force from the interests of a single person, an intra-personal conflict. The other conflict is between reasons which derive their force from the interests of multiple persons, an inter-personal conflict. Nagel argues that we need different principles for the different kinds of conflicts. Proceeding in the same manner in both cases would overlook the significance of the distinction between persons. This is because treating interests of different persons as if they belonged to one person “distorts the nature of the competing claims”.³¹ It is the distortion of moral claims that individuals can press against one another that explains why inter-personal trade-offs differ crucially from intra-personal trade-offs.

Moral claims matter because persons have the ability to understand, evaluate and respond to reasons. Persons differ thereby from other conscious animals. They can take into consideration and act on reasons that other persons give them. One can act towards animals in ways that are justifiable or unjustifiable. But only towards persons can one act in ways that are justifiable or unjustifiable to these persons. This indicates a difference in the way animals and persons can be thought to be members of a moral community. Animals can be passive members insofar as moral norms can be about them; persons can be active, self-legislating members. Because persons can understand and respond to reasons, we can stand in a special relation with persons of acting in ways that are justifiable to them. This means that persons not only have a first-person standpoint, but also a second-person standpoint, i.e. the ability and the moral standing to direct claims towards others. Because they have such moral standing, ignoring their claims would disrespect them and their special moral status.³²

In addition to emphasizing the importance of claims, Nagel also criticizes Rawls’s positive proposal regarding how to respect the separateness of persons. He objects that Rawls’s choice behind the veil of ignorance may allow individuals to simply balance different lives against one another as if they were *merely possible* lives

³¹ Nagel, *The Possibility of Altruism*, p. 138.

³² Similar thoughts are discussed by Scanlon and Darwall. See Scanlon, *What We Owe to Each Other* (Cambridge, MA.: The Belknap Press of Harvard University Press, 1998), pp. 103-7, ch. 4; and Stephen Darwall, *The Second-Person Standpoint* (Cambridge, MA.: Harvard University Press, 2006), esp. chs. 1-2, 6, 12. Scanlon calls this relation “mutual recognition” (p. 162).

of a single agent. Yet bad lives are not mere possibilities, they are actual lives of actual and distinct persons.³³ In Chapter 2 (*Separate Persons Behind the Veil*), I take up this criticism of Rawls's veil of ignorance and explore further what this argument means for veil of ignorance arguments in general. Nagel, however, not only criticizes Rawls's proposal but also develops his own positive proposal. Rather than conflating different moral claims, Nagel argues that we should accept that our moral concern is stratified between different loci.³⁴ Morality includes, for Nagel, a form of impartial concern towards each person. This is achieved by placing oneself into the shoes of everyone else. This allows us to see the moral perspective of others. Our concern for others will remain fundamentally fragmented, however. It is one-by-one. For this reason, Nagel proposes that we should strive for a form of unanimity. Our actions should be justifiable and acceptable to everyone. This would respect the nature of competing claims. To determine whether actions are justifiable, Nagel considers the idea that we should choose the action which is least unacceptable to the person to whom it is most unacceptable.³⁵ Nagel's form of the separateness of persons objection thereby already contains the seeds for a positive proposal of a broadly contractualist morality. I elaborate on Nagel's proposal for how to respect each person's claims separately in Chapter 5 (*Aggregation, Balancing, and Respect for the Claims of Individuals*).

³³ Nagel, *The Possibility of Altruism*, p. 140.

³⁴ In *The Possibility of Altruism* (at pp. 141-42), Nagel considers a solution in which the chooser expects to lead all lives as separate lives. In effect, Nagel imagines here a decision-maker who needs to decide trade-offs of various post-fission selves to whom the decision-maker is equally concerned. In the main text I focus on Nagel's later proposal that he develops in *Mortal Questions*, ch. 8; and Thomas Nagel, *Equality and Partiality* (Oxford: Oxford University Press, 1991): chs. 4-7

³⁵ Nagel doubts whether this is a complete solution to the problem. See Nagel, *Mortal Questions*, pp. 122-25 and Nagel, *Equality and Partiality*, chs. 4, 7. Nagel also refers to a similar idea present in Scanlon's contractualism, see T.M. Scanlon, "Contractualism and utilitarianism," in *Utilitarianism and beyond*, ed. Amartya Sen and Bernard Williams (Cambridge: Cambridge University Press, 1982), pp. 102-28; and Scanlon, *What We Owe to Each Other*.

IV. Two Versions of the Separateness of Persons Objection

Thus far, I have discussed reasons why the difference between intra-personal and inter-personal trade-offs matters. Consider now the following remarks that Rawls makes about individual rationality:

“[Each] man in realizing his own interests is certainly free to balance his own losses against his own gains. We may impose a sacrifice on ourselves now for the sake of a greater advantage later. A person quite properly acts ... to achieve his own greatest good, to advance his rational ends as far as possible.”³⁶

Then Rawls asks in the next sentence: “[Why] should not a society act on precisely the same principle applied to the group”. He strengthens this case by introducing the impartial spectator. The impartial spectator imagines herself to be in everyone’s position and then chooses principles of justice. Any advantages in one position are cancelled out, for the impartial spectator, by disadvantages in another position. This means in Rawls’s words:

“This view of social cooperation is the consequence of extending to society the principle of choice for one man, and then, to make this extension work, conflating all persons into one through the imaginative acts of the impartial sympathetic spectator. Utilitarianism does not take seriously the distinction between persons.”³⁷

One source of Rawls’s objection is, therefore, that utilitarianism pretends that all persons belong to one system of desires. This is particularly evident in the case of the Rawls’s version of the impartial spectator. The device of the impartial spectator fuses all persons into one person. It can achieve impartiality only by treating all person’s lives as one life the spectator will lead.³⁸ A second form of objection is also implicit in the last quote. Principles of individual choice are appropriate for choices for single persons, but they are inappropriate for situations of mutually

³⁶ Rawls, *A Theory of Justice*, rev. edn., p. 21.

³⁷ Rawls, *A Theory of Justice*, rev. edn., p. 24.

³⁸ Rawls, *A Theory of Justice*, rev. edn., p. 161-66.

advantageous social cooperation. The right principle of regulation depends on what is supposed to be regulated.

The two parts of the objection represent two different versions of the separateness of persons objection. The different versions of the objection in turn indicate two different requirements for moral theories. The first part is concerned with the failure of utilitarian philosophers to respect the separateness of persons in the arguments they advance. Utilitarianism fails insofar as the justification given for it does not respect the separateness of persons. I will call this the *justificatory requirement*. The second part is independent from the specific reasons given for the act utilitarian principle. By stating that the utilitarian principle is not suitable in its application to problems of social regulation, social cooperation, and justice, Rawls indicts act utilitarianism as a criterion of rightness. This part is concerned with utilitarianism's failure to respect the separateness of persons in the deontic verdicts it gives. I will call this the *substantive requirement*.

The difference between the two requirements is the following. The justificatory requirement is less sweeping in its implications. The idea is that moral reasons or arguments can violate the separateness of persons. Strictly speaking this criticism does not say that all forms of act utilitarianism violate the separateness of persons, but rather that the reasons given for act utilitarianism can violate the separateness of persons. If there is an alternative justification compliant with the separateness of persons, then act utilitarianism may turn out to be the correct position. By contrast, the substantive requirement on the other side rules out act utilitarianism regardless the reasons given for it. To make the contrast clear, under a substantive version the separateness of persons functions as a constraint on first order moral principles, under a justificatory version the separateness of persons functions as a constraint on moral arguments on behalf of these principles.³⁹

³⁹ A complication arises when we consider rule utilitarianism (or other two-level moral theories). For example, in *On What Matters* Parfit provides a consequentialist theory which may avoid violating either constraint that the separateness of persons imposes. Parfit's rule consequentialism is justified by appeal to the contractualist idea that it is a principle everyone can accept, *seemingly* respecting the justificatory requirement. His rule consequentialism approximates a form of common sense morality that includes, for example, legitimate partiality as the best way to maximize the good in the long-run, *seemingly* respecting the

To see how the two requirements differ in practice, consider the use of the separateness of persons in Henry Sidgwick's *Methods of Ethics* (1874). Sidgwick considers the argument that accepting rational egoism, the theory that one should aim to maximize one's own good, should lead one to accept utilitarianism, the theory that one should aim to maximize the universal good. The egoist accepts that we should sacrifice one's present happiness for one's future happiness. Why should this not lead one to sacrifice one's own happiness for someone else's happiness? Sidgwick rejects this reasoning. He argues that the distinction between persons is "taken as fundamental in determining the ultimate end of rational action for an individual".⁴⁰ We cannot simply reason from rational egoism to utilitarianism. The separateness of persons forbids this. Sidgwick is nonetheless a utilitarian. While he believes in the importance of the separateness of persons, he does not believe that it undermines utilitarianism. After the quoted passage, Sidgwick proceeds to argue that the justification for utilitarianism does not rely on an extension of rational choice according to rational egoism to choices between persons. This indicates that Sidgwick understands the separateness of persons only as a justificatory demand and not as a substantive demand.⁴¹

substantive requirement in its deontic verdicts. See Derek Parfit, *On What Matters*, vol. 1 (Oxford: Oxford University Press, 2011), chs. 16-17. I discuss how to interpret gray areas like these with respect to the justificatory requirement in Chapter 2 (*Separate Persons Behind the Veil*). I leave open how to understand the substantive requirement in cases like these.

⁴⁰ Henry Sidgwick, *The Methods of Ethics* (Chicago: University of Chicago Press, 1962), pp. 418-19, 498. Quote at p. 498.

⁴¹ What complicates this interpretation of Sidgwick is that he, himself, provides an argument for utilitarianism that seems to fail the justificatory demand. Sidgwick argues by analogy in favor of utilitarianism. He first establishes that there is no reason to discount good for being merely in the future. He then provides an argument that just as the individual good is composed of different goods at different points in time, the "universal good" is composed of the different goods of different individuals. Sidgwick concludes: "I obtain the self-evident principle that the good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other". Sidgwick, *The Methods of Ethics*, pp. 380-82, quote at p. 382. The context makes clear that Sidgwick's principle of the "universal good" is his utilitarian theory of morality. The passage provides us with a dilemma. On one reading, Sidgwick's use of the analogy is merely an analogy that helps understanding the utilitarian principle. The analogy is, however, not part of the argument for utilitarianism given that utilitarianism is, for Sidgwick, self-evident and needs no argument. On the other reading, Sidgwick uses the analogy in his argument, in which case he violates the justificatory demand of the separateness of persons. That Sidgwick's argument may violate something like the separateness of persons has been observed even before Gauthier's and Rawls's objections. See

Sidgwick would need a reason for why the separateness of persons is only a demand on moral justification and not on first-order moral principles or theories. David Brink suggests a possible reason.⁴² Brink suggests that the separateness of persons is a substantive requirement only for theories of rationality and not also for theories of morality. According to Brink, the core of the separateness of persons is a principle about uncompensated sacrifice. The separateness of persons tells us that no sacrifice can be imposed on an agent without compensation for it. A theory of rationality that is time-neutral fulfills this. All sacrifices to an individual at certain points in time in her life are compensated by benefits to that same individual at other points in her lifetime. Yet Brink thinks that while this principle is plausible as a principle of rationality, it is implausible as a principle of morality. The reason for this is the skeptical reason we encountered earlier that requiring compensation to an individual for every sacrifice to her would lead to a Paretian morality with no duties to aid others.⁴³ As I indicated above, I do not believe that the best interpretation of the separateness of persons is one that sets such stringent limits on aggregation such that it is ruled out by any sound moral theory. My chapter 5 on aggregation (*Aggregation, Balancing, and Respect for the Claims of Individuals*) lays out the role the separateness of persons can play in thinking about inter-personal aggregation without having extreme and unjustifiable implications.

V. The Justificatory Requirement

Remember that the justificatory requirement of the separateness of persons is the requirement that arguments for a first-order moral principles must respect the separateness of persons. How can a justification fail to respect the separateness of

A.R. Lacey, "Sidgwick's Ethical Maxims," *Philosophy* 34 (1959): 217-28, at p. 219; and Geoffrey Russell Grice, *The Grounds of Moral Judgement* (Cambridge: Cambridge University Press, 1967), pp. 195-97.

⁴² David O. Brink, "Sidgwick and the Rationale for Rational Egoism," in *Essays on Henry Sidgwick*, ed. Bart Schultz (Cambridge: Cambridge University Press, 1992), pp. 199-240, at pp. 207-15.

⁴³ Brink, "The Separateness of Persons, Distributive Norms, and Moral Theory".

persons? I will first highlight three different ways that are each modelled on the choice between alternative principles of justice.

The first way is to choose by imagining each person to be part of the chooser *at once*. This is the crudest way to violate the separateness of persons. Here each person is imagined to be only part of one system of desires and ends. The interests, desires, moral claims and so on of each person are integrated into one system by assuming that they all form part of this system. This argument considers a society as one social entity, or humanity as one super-person. This paradigm case of violating the separateness of persons is seldom expressly defended. J.J.C. Smart comes close to embracing it once. Given that it is rational for us to go to the dentist in order to avoid the pain of a toothache, he asks why should we then not impose pains akin to a dentist visit on others to avoid pains akin to a toothache.⁴⁴ The question at least suggests a model of decision-making where all pains are balanced out as if they belonged to one life.

Smart's argument could, however, also be interpreted in a second way. This second way is to imagine to be in each person's position *in turn*. The clearest exposition of this can be found with C.I. Lewis, who argues that the correct way to assess value is to imagine to be in each person's position in *seriatim*.⁴⁵ We can justify utilitarianism then by the following argument. Imagining to be in each position allows us to compare and cancel out positive and negative experiences that we had at different points in time. The overall evaluation of the process of imagining oneself to be in each position will therefore be the net sum of positive minus negative experiences. One natural way to understand the suggestion of sympathetic imagination is that it aggregates all lives into one long life.

The third way is to imagine each person's life *as a possibility* of one's own future life or as a possibility of one's own actual life. In the latter case, the decision-maker would be temporarily ignorant of their distinctive features and would have to

⁴⁴ J.J.C. Smart, *An Outline of a System of Utilitarian Ethics* (London and New York: Cambridge University Press, 1961), p. 26.

⁴⁵ Clarence Irving Lewis, *An Analysis of Knowledge and Valuation* (La Salle: The Open Court Publishing Company, 1946), pp. 546-47. A similar statement is made in R.M. Hare, *Freedom and Reason* (Oxford: Clarendon Press, 1963), p. 123.

choose a distribution of benefits and burdens not knowing which life she is leading. While the second argument supported total utilitarianism, this argument rather seems to support average utilitarianism. The idea is that we should choose to maximize our own expected utility if we could choose between different social arrangements. In the absence of knowledge of our position in society we would assume equal probabilities for each position. Our choice will then coincide with the highest average utility level in line with what average utilitarianism dictates.⁴⁶ I will discuss both whether this use of the veil of ignorance argument is coherent and whether it really supports average utilitarianism in Chapter 2 (*Separate Persons Behind the Veil*). Even if it was coherent, this method overlooks the separateness of persons by turning actual lives into merely possible lives of a single person. And it makes a significant moral difference whether a bad life is a possibility that gets written off if it does not materialize, or whether someone has to live a bad life no matter what.⁴⁷

VI. The Substantive Requirement

Just as there is not only one way in which the justificatory requirement of the separateness of persons can be breached, there is not only one substantive requirement of the separateness of persons. I group the requirements under four headings.

First, a theory can be charged with overlooking the *separateness of persons* in a narrow sense. Any theory that is fully aggregative overlooks the importance of respecting the standpoints of individuals one-by-one. Simple aggregation across individuals cannot reflect the importance of different standpoints. For example, a principle that would license the imposition of significant burdens on single

⁴⁶ A version of this argument might support total utilitarianism. We need the following additional assumptions: (1) The choice can be made over variable population sizes so that the chooser is not guaranteed to exist, (2) Comparativism between life and death is true. In calculating the expected utility for this gamble, we need to assign a welfare level of zero to those possibilities where we do not exist. But average utilitarianism defined as the principle that selects the highest average welfare of those who exist disregards those who do not exist even under the assumption of comparativism.

⁴⁷ See Nagel, *The Possibility of Altruism*, pp. 138-39.

individuals for minor benefits to very many other individuals, cannot plausibly be said to respect the standpoints of individuals one-by-one. This means that the requirement to respect all standpoints one-by-one, itself an idea about moral justification, gives rise to a requirement about the substantive content of moral principles. This is because no interpretation of the justificatory demand could license a principle that allowed such aggregation.

A second example of this violation of the separateness of persons is treating many small harms occurring to different individuals as equivalent to the same harms occurring all to one person. It would be wrong to think that a short, minor pain experienced by 50 people is the same as pain of 50 times the intensity or duration. The reason for this is that there is no pain of such higher intensity or duration for anyone. There is no standpoint from which the imposition of the pain is as bad. This aspect of the separateness of persons in a narrow sense applies also to some non-persons. Conscious animals which are not persons can have a point of view. In the most basic understanding a point of view just means having subjective experiences such that there is something that it is like to be this entity.⁴⁸ This means that the separateness of such animals matters to some extent. For example, imagine we could kill one cow in order to marginally increase the comfort of many cows in the herd. It does not seem plausible to me that we are permitted to kill the single cow to produce trivial dispersed benefits to many cows.⁴⁹ But not all aspects of the separateness of persons also carry over to the separateness of cows.

One example of this involves a different objection to utilitarianism. Not only its aggregative structure, but also its welfarist commitment is in violation of the separateness of persons. By basing one's judgment of different states of affairs solely on their respective distribution of welfare, utilitarianism overlooks relevant non-welfare factors. The most important one that I have mentioned already are the moral claims individuals can raise. To recognize the importance of different standpoints and to recognize their second-personal authority means that we have to take

⁴⁸ See Godfrey-Smith, *Other Minds*, ch. 4. The idea of understanding consciousness as "what it's like to be" goes back to Nagel, *Mortal Questions*, ch. 12.

⁴⁹ See also John Halstead, "The Numbers Always Count," *Ethics* 126 (2016): 789-802, at pp. 795-96.

seriously moral claims on our actions. A moral claim expresses second-personal authority in a direct way. For example, for utilitarianism it does not matter how well off individuals could have been, something that is relevant in determining the strength of their claim on others. Similarly, for utilitarianism it does not matter whether there are special moral relations between individuals. It does not ultimately matter whether individuals are trustees of someone's interest, promisors or promisees, friends or family members, are responsible for someone's plight, beneficiaries of sadistic pleasure of the misfortune of others, and so on. Only the vector of welfare levels ultimately matters.⁵⁰ However, accepting the authority of the second-person standpoint means that moral relations are important. Here, again, the demand to respect moral claims in the justification of principles gives rise to a condition about the substantive content of these principle. No welfarist theory can do justice to the fact that separate persons occupy separate standpoints from which moral claims can be raised. This is because a moral theory which focuses only on information about welfare levels is necessarily insensitive to the presence or absence of moral relations.

The second way in which the separateness of persons can be violated is by overlooking the *unity of the individual*. This can be regarded as the flipside of the separateness of persons in a narrow sense. While the separateness of persons sets limits to the kinds of permissible trade-offs, the unity of the individual demands that trade-offs are allowed and sometimes required within a person's life.

While the separateness of persons is built on the idea that there is not one social entity but rather various individuals with their own different first-personal standpoints, the unity of the individual is based on the negative of this idea. In the case of an individual, there is only one entity and no competition between the claims

⁵⁰ I follow here Sen's seminal definition of welfarism according to which welfarism is an informational constraint on moral judgment in which only. He writes that "[if] all the personal-utility information about two states of affairs that can be known is known, then they can be judged without any other information about these states". Amartya Sen, "Utilitarianism and Welfarism," *Journal of Philosophy* 76 (1979): 463-89, quote at p. 461. My formulation rules out even a wider understanding of welfarism according to which well-being is the only fundamental value. If we believe that benefits from sadistic pleasure ought not count, then this is accounted for by bringing in values besides welfare.

of different individuals. This means that prudential justifications can become available in the case of intra-personal trade-offs. In situations of pure trusteeship in which our actions have no effects on third parties, we are confronted only with the claim of that individual. The individual would have no reason to object to a prudential justification. Such a justification looks after that person's interests. Furthermore, by being sensitive to the values of the individual it also does not impede an individual's ability to govern their own life.

The third way in which moral theories can be charged with overlooking the separateness of persons is a combination of the two previous points. This is the *difference between the unity of the individual and the separateness of persons*. The separateness of persons requires us to set limits to permissible trade-offs, the unity of the individual requires us to be lenient when it comes to permissible trade-offs. Taken together this means that a moral theory should be sensitive to the difference between inter-personal and intra-personal trade-offs. This is the most famous illustration of the separateness of persons objection and also the one I began my introduction with.

A fourth and final component of the separateness of persons is the *separateness of agents*. The previous three components were all related to what we can and cannot do to people. They regarded persons as passive recipients of harms and benefits. But as persons we are also agents who act in the world, as opposed, for example, to merely patients, beneficiaries, or victims. The separateness of agents is concerned with the separateness of different agential perspectives.

Some theories, notably utilitarianism, embrace what can be called the doctrine of negative responsibility.⁵¹ This doctrine holds that we are equally responsible for what we fail to do (or fail to prevent others from doing) as we are for what we do. Embracing the doctrine of negative responsibility means that we fail to distinguish between what an agent does and what an agent lets others do. It does not distinguish

⁵¹ See Bernard Williams, "A Critique of Utilitarianism," in *Utilitarianism: For and Against*, ed. J.J.C. Smart and Bernard Williams (Oxford: Oxford University Press, 1973), pp. 77-150, at pp. 93-100.

which contributions belong to particular agents in the causal web. This overlooks the separateness of agents.⁵²

The separateness of agents is best explained through an appeal to the importance of the second-personal standpoint. From the second-personal standpoints, individuals can hold us to account for our actions. The separateness of agents is linked to the other side of this accountability relation. It is linked to the perspective of an agent who has to answer the demands of others on her conduct. Such an agent must be able to answer this call for justification. She must be able to respond that what she *in particular* did was justified. The demands of second-person authority are relational; they hold between specific persons. The doctrine of negative responsibility does not, however, distinguish in this manner between agents. Agents cannot point out what *they* have done.

The separateness of agents can also be relevant to the ability and responsibility to live well. One's responsibility to live well and give meaning to one's life is executed by one's actions and omissions. Our actions determine what projects we pursue and how we want to live our lives. The doctrine of negative responsibility asks us to make no distinction in our moral reasoning between our own actions and projects and those of others. Yet such a distinction is necessary for these projects to play the important meaning-giving role for our own life in particular.⁵³ Negative responsibility thereby does not take seriously the separateness of persons. It fails to respect the demand that each person has their separate life to lead, and their own separate projects that are central to their life.

VII. The Separateness of Persons, Attitudes, and Rightness

In my discussion, I have assumed that respecting the separateness of persons is an important requirement for the deontic judgments of a moral theory. In other

⁵² Bernard Williams, "Persons, Character and Morality," in *The Identities of Persons*, ed. Amélie Oksenberg Rorty (Berkeley: University of California Press, 1976), pp. 197-216, at pp. 200-1; and F.M. Kamm, "Moral Status and Personal Identity: Clones, Embryos, and Future Generations," *Social Philosophy & Policy* 22 (2005): 283-307, at pp. 290-91.

⁵³ See also Williams, "A Critique of Utilitarianism," pp. 108-18.

words, respecting the separateness of persons is a matter of what actions are or are not morally permissible. One challenge to the separateness of persons is that this misunderstands the appeal of the separateness of persons. Richard Yetter Chappell gives such an argument.⁵⁴ His argument departs from the idea that violating the separateness of persons is wrong because it treats individuals as mere value receptacles. Compare this with the following statement by James MacKaye, a utilitarian of the early 20th century, elaborating on the utilitarian idea of justice:

“In a manner very similar to that whereby the engineer in the foregoing example determines the factors upon which depends the maximum production of steam, Justice must seek to determine the factors upon which depends the maximum production of happiness. ... [Just] as a boiler is required to utilize the potential energy of coal in the production of steam, so sentient beings are required to convert the potentiality of happiness resident in a given land area into actual happiness.”⁵⁵

He continues his elaboration of the production of happiness:

“Each human being is, in the first place, in his own person, the immediate sentient agent, the happiness-producing mechanism, in whose sensorium the finished product of all successful human effort – happiness – is finally turned out.”⁵⁶

In MacKaye’s statements persons are producers of happiness and happiness is a product which manifests itself in a person. While MacKaye embraces this idea, Tom Regan takes this to be decisive objection against utilitarianism. He elaborates on his objection by using a metaphor of two cups.⁵⁷ Two cups are filled with a sweet liquid, call it value. We can move the liquid from one cup to the other. All that matters is the liquid. The cups themselves are only receptacles or containers of value. They are entirely interchangeable or fungible. The cups do not matter. Nothing bad happens

⁵⁴ Chappell, “Value Receptacles”.

⁵⁵ James MacKaye, *The Economy of Happiness* (Boston: Little, Brown, and Company, 1906), pp. 190-91.

⁵⁶ MacKaye, *The Economy of Happiness*, p. 196.

⁵⁷ Tom Regan, *The Case for Animal Rights* (Berkeley: University of California Press, 1983), pp. 205-6; and “The Case for Animal Rights,” in *In Defense of Animals*, ed. Peter Singer (New York: Basil Blackwell, 1985), pp. 13-26, at pp. 19-20.

if one breaks and is replaced by another cup. But persons are unlike Regan's cups. They are not mere value receptacles. They matter in themselves.

Chappell then argues that one can accept this idea that the standpoints of individuals matter while also endorsing a moral theory that is extensionally equivalent to classical utilitarianism. He construes the value receptacles objection as an objection that utilitarianism treats individuals as fungible. Money bills, for example, are completely fungible. We can simply replace one money bill with another without losing anything of value. Persons, however, cannot be treated as fungible in the same way. This is because, Chappell argues, persons have final value. They are ends in themselves and do not only contribute instrumentally to the good in the way that money bills do.⁵⁸ This difference should be reflected in our moral attitudes. In a trade-off between two money bills we should be indifferent between which one of the two continues to exist. In a trade-off between two persons we should be torn between which one of the two continues to exist. The attitude of conflict and regret expresses the idea that separate persons are separate ends and fulfills the demand of the separateness of persons, according to Chappell.

Chappell's argument applies not only to persons, but also to any entity that has final value. Sites of natural beauty might have final value; great artworks have final value. Their final value indicates that these entities are not entirely replaceable. The entities are not merely constituents of some overall good but are separate sources of value. We can bring this out, like Chappell, in terms of the attitudes that we should take towards entities with final value. It might be fine to sacrifice one painting by Monet in order to save five equally great paintings by Magritte. Nonetheless, it would be appropriate to feel conflict in this case. There are strong grounds for regret in sacrificing the Monet.⁵⁹

⁵⁸ This need not be due to that entity's intrinsic properties. It may also be due to that entities extrinsic or relational properties. For example, sites of natural beauty may have final value and count in themselves only because they are God's creation. For the distinction between intrinsic value and final value see Christine M. Korsgaard, "Two Distinctions in Goodness," *Philosophical Review* 92 (1983): 169-195.

⁵⁹ The feeling of conflict when trading off entities with final value can also enlighten a conservative attitude that G.A. Cohen has argued in favor of (G.A. Cohen, "Rescuing Conservatism: A Defense of Existing Value," in *Finding Oneself in the Other*, ed. Michael Otsuka (Princeton and Oxford: Princeton University Press, 2013), pp. 143-74).

If this were all that the separateness of persons was saying, then there should be no difference between the separateness of persons and the separateness of artworks. However, persons are importantly different from artworks in a way that explains why the separateness of persons is of greater importance. Persons have a rich first-personal perspective and a second-person authority to raise claims on others. This means that the concern for persons involves not only a recognition of separate sites of final value, but a concern for the moral relationship in which agent and patients stand. No such comparable concern matters for the relationship between an agent and Michelangelo's *Pietà*. Taurek phrases this nicely when he writes that in the case of the loss of an arm of the *Pietà*, we are concerned with the loss *of* something. However, in the case where a person loses an arm, we are concerned with the loss *to* someone.⁶⁰ Our concern is not that something of value disappeared or that value was diminished, but rather that there has been a loss to a person with whom we empathize and with whom we stand in a particular relationship. This concern makes it possible that we can *owe it to* persons to act in certain ways. The separateness of persons is distinct from the separateness of artworks insofar as it is centrally about deontic verdicts and not fitting attitudes. Owing behavior to a being implies a directed duty. The duty is not owed to persons as a whole, but to *particular* and separate persons. This should be reflected in its deontic verdicts.

VIII. The Dissertation

Thus far I argued that the separateness of persons imposes two distinct requirements on moral philosophy. One is a requirement regarding the justification of moral theories, the other is a requirement for the first-order content of moral theories. In the previous section I defended the view that the first-order content – i.e. the verdicts of moral permissibility – matter rather than the moral attitudes which the moral theory demands. Before I turn to the content of what the separateness of persons in my view demands, I address one further objection to the idea that the

⁶⁰ Taurek, "Should the Numbers Count?," pp. 306-8.

separateness of persons is morally relevant. The challenge is that the separateness of persons relies on mistaken views about personal identity. Derek Parfit has argued for a revisionist view of personal identity according to which our personal identity – i.e. whether we persist as the same person – is not what matters. Instead, what should matter to us in thinking about prudential or anticipatory concern is Relation R, a relation of psychological connections with a future self.⁶¹ In Chapter 1 (*Why It Does Not Matter What Matters*) I take on this challenge. I argue that even if we grant Parfit's views on the metaphysics of personal identity and on "what matters" for prudential and anticipatory concern, it does not follow that the separateness of persons is unimportant.

In the remainder of my dissertation, I further develop my view of what a moral theory that respects the separateness of persons requires. I have already expanded on both requirements and argued that the separateness of persons is morally important because of the importance of each person's first-person and second-person standpoint.

I begin with a discussion of the link between the justificatory requirement of the separateness of persons and the veil of ignorance as a thought experiment. In Chapter 2 (*Separate Persons Behind the Veil*), I argue that prominent examples of the veil of ignorance, John Harsanyi's and John Rawls's, fail the justificatory requirement. I argue, however, that Ronald Dworkin's veil of ignorance meets this requirement and highlight what is different about his veil. In Chapter 2 I also address a question of the substantive requirement of the separateness of persons in the context of distributive justice. I argue that there are two ways in which principles chosen behind a veil can respect the separateness of persons. One way is Rawls's which restricts the principles to the basic structure of society. The other way is Dworkin's where individual choices behind the veil only influence an entire hypothetical insurance market. Trade-offs are, in Dworkin's theory, determined by an interplay of various individual decisions and not by these decisions themselves. My discussion here focuses on the difference between the unity of the individual and the separateness of persons.

⁶¹ Parfit, *Reasons and Persons*, pt. 3.

As I explained earlier, one component of the unity of the individual and the separateness of persons is the separateness of persons in a narrow sense. The separateness of persons in a narrow sense comes out most clearly in discussions of aggregation. As I have argued, there is an important connection to interpersonal moral theories which place emphasis on the importance of moral relations. Three chapters are devoted to this question.

I start with a challenge for contractualist moral theories. Contractualism, as proposed by T.M. Scanlon, is perhaps the best developed interpersonal moral theory on offer. The challenge I address is how contractualism should address cases in which risks of harm and benefit rather than certain harms and benefits are at stake. In Chapter 3 (*Contractualism, Complaints, and Risk*), I argue against both traditional ex ante contractualism and ex post contractualism. Neither view distinguishes between different kinds of risk. I argue that distinguishing between objective and epistemic risk opens up the possibility for a third view that I call objective ex ante contractualism. This view, I argue, provides us with the best model of justifiability to each and provides us with a plausible model for addressing impositions of risk.

I supplement my argument in Chapter 4 (*Skepticism about Aggregation and Uncertain Rescues*) by considering alternative versions of ex post contractualism which I do not consider in the previous chapter. None of the versions I consider in this chapter are superior in avoiding implausible forms of inter-personal aggregation to the interpretation of ex post contractualism that I argue against in Chapter 3. My discussion in Chapter 4 thereby bridges my discussion of risk and uncertainty with my discussion of aggregation in the following chapter.

Chapter 5 (*Aggregation, Balancing, and Respect for the Claims of Individuals*) is devoted to questions of aggregation. I discuss the problem of how to reconcile anti-aggregationist moral theories with intuitive verdicts that are more permissive about aggregation. In particular, I provide a theory that I call Hybrid Balance Relevant Claims. My view is like others a middle ground between a theory that is fully aggregative and theories that rule out all forms of aggregation. I accept that sometimes a great number of weaker, yet relevant claims can outweigh single stronger claims. I develop my theory by drawing a contrast between two different

kinds of intermediate positions, one with greater affinity to aggregation and my own theory with lesser affinity to aggregation.

In Chapter 6 (*Constraints, You, and Your Victims*), I turn to the importance of the separateness of agents. The chapter discusses the morality of harming and in particular the paradox of deontology – i.e. the seeming paradox that although all rights violations matter equally, we should not violate a single right to prevent a large number of comparable rights violations. I argue for a new relational agent-based justification for deontological constraints. The justification is based on the special relation between the agent and her victims. This justification relates to the separateness of agents. It is the relational feature of *you* harming someone in particular that matters. I argue that my relational agent-based justification can explain why we are not permitted to minimize our own rights violations. I also point out how my relational justification can avoid the charge of being unduly self-concerned.

My dissertation thereby engages in a wide range of topics. I discuss personal identity, distributive justice, egalitarianism, contractualism, inter-personal aggregation, the morality of saving from harm, the morality of harming, and individual rights. My discussion is not exhaustive of the topics that the separateness of persons has been taken to be important for. For example, I do not discuss either libertarianism or liberalism in any depth.⁶² What should we say about this diversity in topics allegedly related to the separateness of persons? A pessimistic conclusion is that the separateness of persons is an elusive concept. When properly analyzed the different references to the separateness of persons turn out to be different arguments which bear no relation to one another. There is no real moral content to the separateness of persons.⁶³ My dissertation as a whole argues for an optimistic

⁶² For this see Martha C. Nussbaum, *Sex and Social Justice* (New York: Oxford University Press, 1999), pp. 61-67; Anthony Simon Laden, "Taking the Distinction between Persons Seriously," *Journal of Moral Philosophy* 1 (2004): 277-92; Matt Zwolinski, "The Separateness of Persons and Liberal Theory," *Journal of Value Inquiry* 42 (2008): 147-65; and Jason Tyndal, "The Separateness of Persons: A Moral Basis for a Public Justification Requirement," *Journal of Value Inquiry* 51 (2017): 491-505.

⁶³ See e.g. Shlomi Segall, "Sufficientarianism and the Separateness of Persons," *Philosophical Quarterly* 69 (2019): 142-55.

conclusion. The demands of the separateness of persons are neither empty nor implausibly stringent. Instead, the separateness of persons is a unifying feature of my discussion of various separate areas of morality. The unity of my dissertation goes along with the separateness of its chapters.

Chapter 1.

Why It Does Not Matter What Matters:

Relation R, Personal Identity, and Moral Theory

I. Introduction

Derek Parfit famously argued that personal identity is not what matters for prudential concern about the future. Instead, he argues what matters is Relation R, a combination of psychological connectedness and continuity with any cause. This revisionary conclusion, Parfit argued, has profound implications for moral theory. It should lead us, among other things, to deny the importance of the separateness of persons to morality. Instead, we should adopt impersonal consequentialism. In this chapter, I argue that Parfit is mistaken about this last step. His revisionary arguments about personal identity and rationality have no implications for moral theory. The importance the separateness of persons has for morality does not turn on whether personal identity rather than Relation R is what matters for prudential concern.¹

¹ For Parfit's views see Derek Parfit, "Personal Identity," *Philosophical Review* 80 (1971): 3-27; "On 'The Importance of Self-Identity'," *Journal of Philosophy* 68 (1971): 683-90; "Later selves and moral principles," in *Philosophy and Personal Relations*, ed. Alan Montefiore (London: Routledge, 1973), pp. 137-69; "Lewis, Parry, and What Matters," in Rorty, *The Identities of Persons*, pp. 91-108; *Reasons and Persons*, pt. 3; and "The Unimportance of Identity," in *Identity*, ed. Henry Harris (Oxford: Oxford University Press, 1995), pp. 13-45. For views similar to Parfit's see John Perry, "The Importance of Being Identical," in Rorty, *The Identities of Persons*, pp. 67-90; Sydney Shoemaker, "Personal Identity: A materialist's account," in *Personal Identity*, ed. Sydney Shoemaker and Richard Swinburne (Oxford: Basil Blackwell, 1984), pp. 67-132; and Jennifer Whiting, "Friends and Future Selves," *Philosophical Review* 95 (1986): 547-80. For a variety of arguments that personal identity is what matters see David Lewis, "Survival and Identity," in Rorty, *The Identities of Persons*, pp. 17-40; Vinit Haksar, *Equality, Liberty, and Perfectionism* (Oxford: Oxford University Press, 1979), pp. 106-13; Vinit Haksar, *Indivisible selves and moral practice* (Edinburgh: Edinburgh University Press, 1991), pp. 158-215; Susan Wolf, "Self-Interest and Interest in Selves," *Ethics* 96 (1986): 704-20; Robert Merrihew Adams, "Should Ethics be More Impersonal? A Critical Note of Derek Parfit, *Reasons and Persons*," *Philosophical Review* 98 (1989): 438-84; Christine M. Korsgaard, "Personal Identity and the Unity of Agency: A Kantian Response to Parfit," *Philosophy & Public Affairs* 18 (1989): 101-32; Mark Johnston, "Reasons and Reductionism," *Philosophical Review* 101 (1992): 589-618; Mark Johnston, "Human Concerns without Superlative Selves," in *Reading Parfit*, ed. Jonathan Dancy (Oxford: Blackwell, 1997), pp. 149-79; Marya Schechtman, *The Constitution of Selves* (Ithaca: Cornell University Press, 1996), ch. 4; and Tim Christie, "Natural Separateness: Why Parfit's Reductionist Account of Persons Fails to Support Consequentialism," *Journal of Moral Philosophy* 6 (2009): 178-95. Otsuka argues that personal identity is a sufficient condition for

When spelling out the moral implications of his view on personal identity and what matters, Parfit mentions a variety of examples. The examples range from revising our views on paternalism and autonomy, abortion, promises and commitments, retribution and desert, and the importance of equality to the separateness of persons objection to utilitarianism. My discussion will be focused on the importance of the separateness of persons objection. As I explained in the introduction to this dissertation, the separateness of persons objection occupies a central place in non-consequentialist moral thinking. Utilitarianism ignores the separateness of persons, the argument holds, because it aggregates all benefits and burdens across different persons. Sometimes, however, we are allowed to aggregate different benefits and burdens. In particular, we are allowed to aggregate when these benefits and burdens fall within one life. This is explained by the unity of the individual. Together the separateness of persons and the unity of the individual demand that we should treat inter-personal trade-offs differently from intra-personal trade-offs. Utilitarianism cannot do this.

Parfit's revisionary arguments concerning morality can be reconstructed as attacking both components of the separateness of persons objection to utilitarianism. One argument holds that Parfit's views on personal identity and what matters undermine the unity of the individual. I examine and reject this argument in Section II. Another argument holds that Parfit's views on personal identity and what matters undermine the separateness of persons. I examine and reject this argument in Section III. A last argument holds that his views render the unity of the individual and the separateness of persons less relevant. I examine and reject this argument in Section IV.

Throughout this paper my strategy is to accept Parfit's arguments concerning personal identity and rationality, and to reject the link he draws from metaphysics and rationality to morality. My strategy thereby differs from what Mark Johnston has

prudential concern (Michael Otsuka, "Personal Identity, Substantial Change, and the Significance of Becoming," *Erkenntnis* 83 (2018): 1229-43). Unger and McMahan propose views which qualify Relation R with a physical realizer. This makes their views in practice close to personal identity. See Peter Unger, *Identity, Consciousness and Value* (New York: Oxford University Press, 1990), chs. 4-5, 7; and Jeff McMahan, *The Ethics of Killing* (Oxford: Oxford University Press, 2002), pp. 66-82.

called “minimalism”. Johnston remarks that many of our practices, like those of morality and rationality, lend themselves to certain metaphysical views. Minimalism then holds that the justification of these practices is independent from the truth of the metaphysical position. Metaphysical positions, like those about personal identity, are epiphenomenal to practices like rationality and morality.² Unlike Johnston, I believe that the truth of metaphysical positions can have an impact on normative practices like morality. Indeed, I accept, at least for the sake of argument, that Parfit is correct about his link between metaphysics and rationality. I only deny that Parfit’s revisionary argument for morality stands.

II. First Argument: Less United Individuals

So why should Parfit’s conclusion about the metaphysics of personal identity have any impact on morality? The first suggestion is that Parfit’s claim that what matters is Relation R rather than personal identity demonstrates that the unity of the individual is unimportant. When discussing the diminished importance of the separateness of persons, Parfit writes: “If the unity of a life is less deep, it is more plausible to claim that this unity is not what justifies maximization”.³ In rational decision-making we are allowed to pursue what will bring about the highest sum-total of well-being. If Parfit is right in holding that the unity of the individual is less important, then this cannot be justified by appealing to the unity of the individual. Maximizing the sum-total of well-being would then seem to be justified differently and apply also in inter-personal trade-off, in line with what utilitarianism demands.

One way to explain Parfit’s claim that the unity of the individual is undermined, is by appealing to what we may call the relevant units of moral and prudential concern. Some nationalists claim that nations have moral importance over and above individuals. Nations matter for their own sake. A nation is, under such a view, a unit of moral and prudential concern. For “moral individualists”, the unit of

² Johnston, “Reasons and Reductionism”; and Johnston, “Human Concerns without Superlative Selves”. Similar arguments are made by Wolf, “Self-Interest and Interest in Selves”; Adams, “Should Ethics be More Impersonal?”; and Christie, “Natural Separateness”.

³ Parfit, *Reasons and Persons*, pp. 334-35.

moral concern is a person's entire life. But other proposals are possible. We could focus on parts of lives (I shall call these "person stages"), or we could focus on time-slice persons, instances in an individual's life without much temporal extension.⁴ Parfit at some point writes that following his view we should "regard the rough subdivisions within lives as, in certain ways, like the divisions between lives".⁵ Elsewhere he speaks of the "partial disintegration" of persons.⁶ This seems to suggest that Parfit regards parts of lives, in particular those with high degrees of psychological connectedness, as the basic units of moral and prudential concern. The idea is that the psychological connections that contain what matters come in degrees. Some of these connections wither away over time. We are more strongly connected to our past and future person stages close in time. While this is a statement about the unit of prudential concern, Parfit's statements indicate that he thinks that the unit of moral concern coincides with the unit of prudential concern. Indeed, his statement about treating rough subdivisions within lives like divisions between lives is made in the context of moral principles.

Moral theories should therefore take person stages, rather than full lives as their objects of principal concern. For example, questions of distributive justice would then arise between person stages rather than between persons. This means that principles of distributive justice would need to be given a greater scope. They would also extend to trade-offs within a person's life, namely to those between one person stage and its future successive person stage.⁷ But since principles of distributive justice would then apply to such a variety of cases, we might think that we have less reason to care about distributive justice. Our intuition that distribution matters is less strong for intra-personal, inter-stage trade-offs. Yet it is unclear why this is the conclusion we should draw from the widening of the scope of principles of justice.

⁴ For similar distinctions see David O. Brink, "Rational Egoism and the Separateness of Persons," in Dancy, *Reading Parfit*, pp. 96-134, at pp. 110-16; and David W. Shoemaker, "Selves and Moral Units," *Pacific Philosophical Quarterly* 80 (1999): 391-419, at pp. 391-92. Shoemaker, following Parfit, calls these persons in parts of their lives 'selves', Brink calls them 'person segments'.

⁵ Parfit, *Reasons and Persons*, pp. 333-34.

⁶ Parfit, *Reasons and Persons*, pp. 335-36.

⁷ Cf. Parfit, *Reasons and Persons*, pp. 332-34.

Instead of revising our intuition about the importance of distributions, we could revise our intuition about intra-personal, inter-stage trade-offs. Perhaps we trust this intuition less since it might derive from traditional views about personal identity that Parfit rejects. If we take this answer to the problem, then we would revise our view on individual rationality. We would no longer be justified to pursue the maximum benefit when facing trade-offs that only affect our life. Principles of distributive justice would be extended to all trade-offs involving different person stages.⁸

This looks like a stand-off between two different ways to adjust our intuitions. However, the defender of Parfit's view has another argument in hand. Talking about person stages is only a useful heuristic. Person stages are united by greater psychological connections and what matters are these connections (Relation R). Introducing person stages can help us to avoid talking about Relation R directly, but it is an imperfect heuristic. The boundaries between different stages are fuzzy and different stages overlap. Once we see this, it is less plausible to just apply our moral principles to different units. To substitute one unit of moral concern for another overlooks the fuzziness around the borders of these moral units. There does not exist a unity of a person stage that is comparable to the unity a defender of the unity of the individual has in mind. We can then rightly ask why we should attach such great importance to the difference between different units when the units are only useful heuristics to refer to persons in different stages of their lives.

But why should we follow Parfit in believing that Relation R leads us to accept person stages as the unit of moral concern? Why should we believe that entire lives are not strongly integrated? Broadly speaking there are two possible arguments, contra Parfit, that lives are strongly integrated. One argument is that something other than Relation R unifies persons. Consider, for example, Kantian replies to Parfit's claims.⁹ These replies can admit that persons are neither metaphysically united nor united via Relation R. Instead, there is something else that unites persons. The

⁸ Cf. Nagel, *Mortal Questions*, pp. 124-25fn16; and Dennis McKerlie, "Egalitarianism and the Difference Between Interpersonal and Intrapersonal Judgments," in *Egalitarianism*, ed. Nils Holtug and Kasper Lippert-Rasmussen (Oxford: Clarendon Press, 2006), pp. 157-73, at pp. 163-67.

⁹ Korsgaard, "Personal Identity and the Unity of Agency"; and Simon Blackburn, "Has Kant Refuted Parfit?," in Dancy, *Reading Parfit*, pp. 180-201.

Kantian response to Parfit highlights that persons are united by the practical perspective and the necessity to act as agents. In a similar vein, David Brink argues that agency is best ascribed to persons rather than person stages. Considering person stages to be agents would lead to an undue proliferation of various, overlapping agents.¹⁰ The second argument, on the contrary, does not introduce any further considerations over and above Relation R that could explain why individuals are unified. Instead, this argument rejects the claim that Parfit's arguments have established that Relation R fades out over time. In line with my general strategy of granting Parfit his claims about metaphysics and rationality, I pursue an argument of the second sort.

Relation R is the relation of psychological connectedness and/or continuity with any cause. Psychological connectedness refers to the *degree* to which the same psychological features are present in two different persons at different times. The psychological connectedness between me, now and me, two seconds ago, is very high. The psychological connectedness between me, now and me, two years ago is lower. I have forgotten some experiences, do not share all of my beliefs, adjusted my plans of life, and so on. Psychological continuity requires a series of overlapping bonds of strong psychological connectedness. Continuity does not require, however, that connectedness is given between earlier and later stages in the series. As such, psychological continuity, unlike psychological connectedness, is a transitive relation.

The idea that Relation R weakens over time requires an interpretation of Relation R in which Relation R comes in degrees. Only then will person stages show a significantly greater extent of R-relatedness than entire lives. I already mentioned that psychological connectedness is a matter of degree. But is psychological continuity as well? Parfit contrasts connectedness as a relation that comes in degrees with continuity indicating that he does not believe that continuity is a relation that comes in degrees.¹¹ Nevertheless, Brink offers a construal of continuity in which continuity is a matter of degree. According to Brink, two persons are more strongly continuous with one another if the individual connections in the chain of

¹⁰ Brink, "Rational Egoism and the Separateness of Persons," pp. 110-16, 121-23.

¹¹ Parfit, *Reasons and Persons*, p. 206.

psychological connectedness that constitutes continuity are stronger.¹² An immediate problem for such a view is that continuity is transitive. Parfit defines continuity as a transitive relation in order for continuity to be a possible criterion of personal identity. Since personal identity is transitive, continuity must be as well.¹³ Continuity is thus defined as transitive precisely to express a form of connection that the non-transitive relation of connectedness does not express. The problem for Brink's view is now that transitivity is defined only as a property of binary relations and not defined for relations that come in degrees.

We can make sense of the suggestion that continuity comes in degrees in another way. We can imagine a family of continuity relations which each specify a different threshold of connectedness that is needed to ensure continuity. A person stage is then more continuous with a past or future person stage if a higher threshold of connectedness is met. For example, continuity_{STRONG} requires that all overlapping chains consist of strong connectedness, continuity_{VERY-STRONG} requires chains of very strong connectedness, continuity_{EXTREMELY-STRONG} requires extremely strong connectedness. Two person stages might then be more continuous if continuity_{EXTREMELY-STRONG} rather than continuity_{STRONG} holds between them.

This construal of continuity is a threshold view. According to this view, the weakest link determines the strength of the degree of continuity of the entire chain. The degree of continuity for an entire life is therefore determined exclusively by the amount of connectedness in the moment where the greatest change occurred. This does not cohere well with the reason for which continuity was introduced. Continuity is distinguished from connectedness to explain the psychological connection between persons over a long period of time. Continuity can explain how an old person is psychologically connected to her childhood person stage. But then it does not appear that it should matter very much how these changes occurred.

Take the example of St Paul who according to the biblical story fell on the road to Damascus, heard the voice of Jesus, and decided to stop his persecution of

¹² Brink, "Rational Egoism and the Separateness of Persons," p. 132fn31; and David O. Brink, "Self-Love and Altruism," *Social Philosophy & Policy* 14 (1997): 122-57, at pp. 138, 141-43.

¹³ Parfit, *Reasons and Persons*, pp. 206-7.

Christians and convert. St Paul's story is one of a single sharp change. Contrast this with a person who lives an erratic life and changes her life's narrative multiple times. Finally, towards the end of her life she, like St Paul, arrives at a point that is very different from her early person stages. As long as none of the individual changes in her life were as drastic as St Paul's conversion, she would, following the current proposal, be more continuous than St Paul. While it may make sense to think that St Paul's life has not been fully continuous, it makes little sense to think that the erratic life has been more continuous than St Paul's. St Paul's life has a clear narrative that is only disrupted by a single incident. No clear narrative exists for the person with the erratic life. Given that continuity is supposed to account for long-term relations, it seems hardly plausible that degrees of continuity should be so sensitive to single points in time. The reason why continuity is distinguished from connectedness as a separate relation is better accounted for by understanding continuity as an on-off relation.

The argument for person stages as the unit of moral concern therefore cannot rely on an analysis of psychological continuity. But I have admitted that psychological connectedness comes in degrees. If we attach primary importance to psychological connectedness, then we can argue that person stages are the relevant unit of moral concern. If, on the other hand, we attach little significance to psychological connectedness, then we have no grounds for believing Parfit's argument that person stages are the relevant unit of moral and prudential concern. In such a case my previous argument has shown that psychological continuity would ensure that we regard entire lives as the proper unit of moral concern.

While for most parts of his argument, Parfit does not distinguish between the two components of Relation R, we can now see that the difference is important. So what is Parfit's argument that psychological connectedness is an important part of what matters? His argument here is very brief. He analyses three components of psychological connectedness/continuity to see whether we care about being connected instead of merely continuous.¹⁴ The first component is memory. If only continuity mattered, then "[i]t should not matter to me that I shall soon have lost all

¹⁴ Parfit, *Reasons and Persons*, p. 301.

of my memories of my past life.” But this is implausible. Indeed, we care heavily about retaining our memories. Also in terms of desires and intentions Parfit claims that we want more than continuity. Our lives should have an overall unity and should not be episodic with continued fluctuations. Thirdly, Parfit claims that there are parts of our character that we do not want to change. Here again, he claims, connectedness matters.

To assess Parfit’s argument, it will be helpful to make the case a bit more concrete. We can take a case where psychological continuity is given but psychological connectedness is low. Alzheimer’s is such a case.¹⁵ A person before the development of Alzheimer’s is psychologically continuous with the person having developed Alzheimer’s. But their psychological connectedness is limited. The person has forgotten many of the memories she once had. It is also likely that many of the intentions or long-term plans that the person had will have changed or she simply will have forgotten them. Maybe there will be further changes due to Alzheimer’s that reduce psychological connectedness. If the person used to be very engaged with intellectual activities, her character will inevitably change when the illness leads to a decline of her reasoning skills. As noted earlier, Parfit has argued that in these cases we do seem to care about our connectedness with these persons. I agree with this to some extent. But I think Parfit relies here on an ambiguity in the locution “what matters”.¹⁶ It matters to us *that* or *whether* these changes happen. The thought of Alzheimer’s is truly frightening to many, including me, and we would strongly want to avoid it.

Yet Parfit needs another claim to support his argument. He needs to say that connectedness matters *once* or *when* these changes happen. In other words, he needs

¹⁵ We can leave aside complications of late stage Alzheimer’s where all psychological connections to one’s previous life are cut and so there is no continuity either. Some authors, notably Jeff McMahan, have held that we have grounds to be rationally concerned with a future Alzheimer’s-Self who is not even psychologically continuous with us (cf. McMahan, *The Ethics of Killing*, p. 65). Here I do not need this stronger claim but only the weaker claim that we have reason to be rationally concerned with a future Alzheimer’s-Self that is continuous with us. Those like McMahan who believe in the stronger claim will also support the weaker claim.

¹⁶ A similar observation about the ambiguity of “what matters” is made by Peter Unger. Unger uses the terms “desirability use” and “prudential use” for the contrast (cf. Unger, *Identity, Consciousness and Value*, pp. 92-97).

the claim that connectedness constitutes the basis for rational self-concern. But here it does not seem plausible to me that we would lose the special bond with the resulting person once we develop Alzheimer's. If someone told us that the person with whom we will be continuously connected will be tortured in the future, we would rightfully be horrified. It would concern and involve us deeply. If we hear that a stranger that is qualitatively similar to us will be tortured, we may have sympathy but will not be as involved as in the previous case. Now how should we react if we hear that a person with whom we will be continuously connected but who will develop Alzheimer's will be tortured? Parfit's claim that connectedness matters should make us be less worried or concerned about this news. We should treat it more like the news of the stranger. Yet I cannot see why we should not react with the same horror and concern to the news as in the case of our continuous self who will not develop Alzheimer's.

Let me illustrate my distinction further with an analogy. Parents often want children to turn out a specific way. At the very least they would like their children to be successful and happy. This matters tremendously to them. But parental love does not relinquish when children do not meet this standard. It does not matter to parents that their child is not successful once this is the way things are. They do not lose the special bond of concern with their children if these happen to be unsuccessful and depressed. Similar things hold in the self-regarding Alzheimer's case. Of course we would prefer a future without Alzheimer's. But this does not mean that we give less importance to our bond with our psychologically continuous Alzheimer's-self.

A second reason to think that psychological connectedness is not what primarily matters is the following. Psychological connectedness will be very high when there is a great overlap in our psychology between past and future selves. But we certainly do not want our life to be static. Even if we are content with ourselves and cannot identify specific parts of ourselves to be changed, we still would want to develop and grow as persons. Parfit to some extent agrees with this general observation, but he remarks that we want our life to have a certain overall unity. The life should not be episodic.¹⁷ But here similar arguments like the ones I raised before

¹⁷ Parfit, *Reasons and Persons*, p. 301.

apply. I can concede that we do not want that our life will be episodic. Such a life would not have the requisite unity or narrative that we strive for. We might even think that such a life could not be a good life. In short, we do not want *that* this happens. But does this also mean we shall lose all special bond or interest in the person who is at the end of our episodic journey through time? I doubt that. Our intuitions about the unity of life are intuitions about what makes a life good, but we will still be concerned with our path through life even if our life is deficient in some sense.

We can make the remarks about change more precise. Some decisions are very likely going to result in psychological changes in the personality of the person making the decision. Take the decision of a young adult from a working-class background whether or not to go to university. If she goes, the would-be student will experience a new social setting very different from the one she is used to. She will be exposed to ideas and avenues radically different from those she would have encountered otherwise. This is confirmed by reports of a culture shock for students from working-class backgrounds in higher education. She can foresee that the university experience will change her. It is foreseeable for the decision-maker that one option will lead to significant psychological changes. The new experience can turn out to be transformative for that person.¹⁸ Psychological connectedness will hold only to a reduced degree between the decision-maker and her future self. This change will only happen, however, if one of the two options is chosen. Assuming that psychological connectedness is the primary part of what matters, this influences the rational assessment of these decisions. The decision to engage in the transformative experience will be less appealing. Any potential gains of higher education will have to be discounted by the fact that we should have less prudential concern for the resulting person. The expected benefits of going to university would need to be very substantial to counteract the lessened concern. This does not strike me as a plausible model for thinking about these kinds of decisions.¹⁹

¹⁸ For a detailed treatment of personally transformative experiences see L.A. Paul, *Transformative Experiences* (Oxford: Oxford University Press, 2014).

¹⁹ For a similar observation see Wolf, "Self-Interest and Interest in Selves," pp. 712-13.

The argument becomes even more pressing in the special case when we regard the change positively as an improvement.²⁰ We do want to change some of our psychological features and would not regard their disappearance as a loss in any way. However, successful improvement of our psychological features would render us less psychologically connected with our past self. If psychological connectedness expresses what matters, then we should have less prudential or anticipatory concern for our successive improved self. In a way this even undermines the rationality of efforts made in order to improve one's character. These efforts are borne out of a concern for an eventually resulting person that will be psychologically less connected with the person having made the sacrifices. If we should have less rational concern for this resulting person, these efforts may not be worthwhile after all.

There is one feature about the improvement argument that might seem problematic. David Shoemaker objects that, contrary to what I have been assuming, cases of improvements do contain a significant degree of connectedness. Most importantly, there is a shared intention of wanting to improve one's character and life. This intention connects these parts of one's life strongly together. The strong connection is evidenced by the fact that we can identify with our past self in a way that we cannot with an even more remote self, like our past self before we made the decision to change our life.²¹

Shoemaker's point is apt for deliberate decisions to improve one's character. But not all improvements need to involve an intention. Earlier, I described decisions which can have a transformative impact on the decision-maker. It seems possible that there are decisions where the decision-maker can foresee that the decision will have a positive transformative impact yet does not choose the option because they intend the improvement. Take the example of parenting. Let us assume that a person

²⁰ This possibility is also discussed by Brink, "Rational Egoism and the Separateness of Persons," pp. 119-21. Christine Korsgaard touches on this issue when she discusses changes that are deliberately brought about by the agent, something she calls "authorial connectedness". Korsgaard, "Personal Identity and the Unity of Agency," pp. 120-23.

²¹ Shoemaker, "Selves and Moral Units," pp. 406-9. For a more extensive discussion of Shoemaker's point on identification see David W. Shoemaker, "Theoretical Persons and Practical Agents," *Philosophy & Public Affairs* 25 (1996): 318-32; at 328-31. He extends on a point made earlier by Parfit ("On 'The Importance of Self-Identity'"). Parfit did not include identification in his discussion of personal identity in *Reasons and Persons*.

foresees that being a parent will induce positive character changes, for example by becoming a more responsible person. But the decision to become a parent may have been made on grounds entirely independent of these changes. In this case, the improvement of the future parent's character is not intended and therefore there is no intention that connects the self of the future parent with the later improved self. Here, too, the fact that the future parent will be less connected to her later self should not make undergoing the improving experience any less rational.

What is present, however, is a second-order desire by the future parent to be a more responsible person. This second-order desire is fulfilled in the case of first-order psychological change while it is frustrated in the case of first-order psychological stagnation. While the second-order connection does hold over time in cases of improvement, many other first-order psychological connections will be weakened. Psychological connectedness can accept cases of improvement only if there is a reason why we should privilege second-order psychological connections over first-order psychological connections.

One suggestion here is related to the idea of self-identification. The idea is that there is a sense of alienation towards those first-order desires that we rather not have while there is a sense of self-identification towards those first-order desires that we wish to retain. Alienation and self-identification do provide us with good reasons for regarding some desires as more properly our own than others. Parfit, when he discusses self-identification, draws a contrast between self-identification and non-identification. Non-identification is marked by an attitude of indifference towards a past self. Indifference in turn is marked by the absence of feelings of pride, shame, regret and the like.²² This analysis of self-identification does not privilege desires that we approve of over those we disapprove of. Shame and regret for having certain desires can just as well provide for self-identification. I think Parfit is right in this construal of self-identification. We talk about people owning up to one's mistakes. A person repentant of a former self that did wrong is not regarding this former self as alien to herself. On the contrary, it would be difficult to understand the intensity of feelings of remorse and guilt if the person would not identify the former self as

²² Parfit, "On 'The Importance of Self-Identity'".

genuinely herself. Of course, sometimes there is a feeling of alienation from our first-order desires. But alienation is not the same as disapproval, the two can diverge. Since this is the case, the importance of self-identification cannot give us a reason why second-order desires are more important psychological connections than first-order desires. This in turn means that improvements do not necessarily ensure psychological connectedness. The reply to the improvement argument fails. It seems then that Parfit's case for psychological connectedness as a central part of the relation of what matters does not stand.

This concludes my discussion of connectedness and continuity. We should interpret Relation R as giving primary weight to psychological continuity as opposed to connectedness. We can retain Parfit's central claim that personal identity is not what matters. What matters instead is Relation R. Psychological continuity can, unlike personal identity, be one to many, as illustrated by fission cases where one brain is divided and inserted into two different brainless bodies. Both resulting persons will then be psychologically continuous with the original person whose brain was divided.²³ But, as it turns out, in our world this difference is not relevant. We do not divide or branch in our real world. For us, personal identity perfectly coincides with psychological continuity. Unlike psychological connectedness this does not come in degree but is an on-off relation. The appropriate unit of moral and prudential concern therefore remains an entire life. The unity of the individual is safeguarded by the unity of psychological continuity.

III. Second Argument: Less Separate Persons

The arguments canvassed so far have sought to undermine the unity of the individual. But instead of focusing on the unity of the individual, we could focus on the separateness of persons. Consider the following famous passage in which Parfit describes his own attitude after coming to believe the reductionist view.

²³ Parfit, *Reasons and Persons*, pp. 254-60.

“There is still a difference between my life and the lives of other people. But the difference is less. Other people are closer. I am less concerned about the rest of my own life, and more concerned about the lives of others”.²⁴

One way to interpret this passage is that Parfit is suggesting that we can have similar relations to other contemporaneous persons as we have to our future selves. This includes Relation R which contains what matters. If the way we are related to our future selves is similar to the way we are related to other distinct people, then this reduces the extent to which we are distinct from other persons. Jennifer Whiting and David Brink have suggested, in a similar vein, that our relation to our future person stages is like the relation to close friends or family.²⁵ If this is so, then we would no longer be justified in putting such great weight on the separateness of persons as a bar to inter-personal aggregation.

There are many different ways individuals can be psychologically related to us. These correspond to the different important features of our mental lives. Sharing memories, intentions, beliefs, or dispositions are ways to be psychologically related. To be one of the psychological relations included in Relation R, the relation has to have a causal component. It is not sufficient that two persons are very much alike in terms of their psychological characteristics. The causal component in Relation R is important to distinguish numerical identity from qualitative identity. Sometimes older people say things like “you remind me of myself when I was young”. This is a statement about qualitative identity. The older person sees many of the features of her own psychology when she was young in the other person. But this psychological resemblance is clearly not sufficient to establish numerical identity.

With regard to causal psychological connections, we should draw a distinction between those connections that are first-personal and those connections which are not first-personal. By first-personal I do not mean that the connections have to be had by the same person. Rather, I understand a first-personal connection as a non-deviant causal connection between first-personal mental states.

²⁴ Parfit, *Reasons and Persons*, p. 281. My emphasis.

²⁵ Whiting, “Friends and Future Selves”; and Brink, “Self-Love and Altruism,” pp. 136-43.

The contrast here is between mental states that are “from the inside” against mental states “from the outside”. It means that connections are presented in the first-personal mode of presentation.²⁶ The connections must be part of a self-centred scheme of one particular point of view. The distinction is best explained with regard to memory. I might have the memory of hearing Parfit speak. The memory is detached from the person Parfit, just as in a dream we sometimes see ourselves from a third-person perspective. This memory is markedly distinct from a memory in which I seem to recall having Parfit’s body and voice and am speaking at All Souls College. This second memory is had “from the inside”. The memory is one in which I occupy Parfit’s self-centred perspective of the world. It is not just that I am imagining how All Souls looked and Parfit’s voice sounded. Rather, it is, in Williams’s words, participation imagery from the perspective that Parfit occupied.²⁷ A second example is the link between an intention and a subsequently carried out action. Intentions entail a first-person perspective; they are intentions that the agent performs an action. Intentions are “inside” of a particular self-centered scheme.²⁸

Memories or intentions of this sort need not presuppose personal identity. Parfit introduces a revision of the concept of memory, originally proposed by Sydney Shoemaker, that he calls quasi-memories.²⁹ In quasi-memories the subject seems to remember an experience from the first-personal point of view, someone had this experience, and there is a non-deviant causal connection between the experience and the memory. Similarly, for someone to have a quasi-intention, one has to have an intention, a subsequent action has to be performed, and the intention must cause the action in the right way.

To see the importance of causation in the case of memory, consider a case in which a person who has been in an accident forgets about her experience. At a later point in time a skilled hypnotist implants imagery of an accident into the minds of

²⁶ Cf. Parfit, *Reasons and Persons*, pp. 220-22.

²⁷ Bernard Williams, *Problems of the Self* (Cambridge: Cambridge University Press, 1973), pp. 43-44; see also J. David Velleman, “Self to Self,” *Philosophical Review* 105 (1996): 39-76, at pp. 48-50.

²⁸ Cf. Velleman, “Self to Self,” p. 70.

²⁹ Parfit, *Reasons and Persons*, pp. 220-22; and Sydney Shoemaker, “Persons and Their Pasts,” *American Philosophical Quarterly* 7 (1970): 269-85.

her audience. By sheer coincidence the hypnotist's imagery of the accident corresponds perfectly to the imagery of the actual accident. Such a case is clearly not one of remembering or quasi-remembering the accident.³⁰ The causal connection is even clearer to see in the case of intentions. What is special about intentions is that intentions can lead directly to actions without agential interference. Intentions are causes of actions.³¹

With these clarifications in mind, the question arises whether psychological continuity and therefore strong psychological connectedness, requires first-personal connections. The first thing to note is that the examples that Parfit gives as elements of Relation R tend to be first-personal connections. For example, when introducing the relations of psychological connectedness and continuity Parfit introduces them after a discussion of quasi-memories and quasi-intentions.³² This gives an indication that Relation R appears to be a plausible criterion for what matters in large part because it contains first-personal connections.

While this is indicative, there are other arguments which strengthen the case for the centrality of first-personal connections. Consider the relation between you, now, and a future person who happens, by fortuitous coincidence, to have the same character, habits and other psychological features as you. In short, the person is qualitatively very similar to you. This relation is not Relation R and does not contain what matters. There are two elements missing in this case. One is the absence of a causal link between your psychology and the future person's psychology. The other element is the absence of first-personal connections. Which one of these two elements explains more satisfactorily why qualitative similarity is insufficient for Relation R?

³⁰ Parfit, *Reasons and Persons*, p. 207. The example is due to C.B. Martin and Max Deutscher, "Remembering," *Philosophical Review* 75 (1966): 161-96, at pp. 174-75.

³¹ Parfit, *Reasons and Persons*, p. 261. In both cases we need further conditions that rule out deviant causal chains. For the case of intentions see Donald Davidson, *Essays on Actions and Events* (Oxford: Oxford University Press, 2001), pp. 74-82; John R. Searle, *Intentionality* (Cambridge: Cambridge University Press, 1983), pp. 83-98; and Alfred R. Mele, *Springs of Action* (Oxford: Oxford University Press, 1992). For the case of memory see Martin and Deutscher, "Remembering," pp. 178-91; and Alan Sidelle, "Parfit on 'the Normal/a Reliable/any Cause of Relation R,'" *Mind* 120 (2011): 735-60, at pp. 744-48.

³² Parfit, *Reasons and Persons*, pp. 204-5.

It is hard to see why causation by itself should make such a big difference. There is no obvious reason why causally sustained psychological connections should be particularly important. There is nothing intrinsic to causation that suitably connects with our concerns of what matters. It is difficult to see how the mere fact that some connections are causally sustained could explain what distinguishes fortuitous psychological connections by accident from ordinary cases of personal identity.

Why causation is important is even more puzzling given that Parfit thinks that any causal link would be sufficient to satisfy psychological connectedness.³³ If what matters is the effect and not how it was caused, why does it matter that it was caused in the first place? One possibility is that something associated with the causal requirement explains why the relation has to be a causal one. In this case then, it would be this extra factor rather than the causal link as such which explains why the relation between the two persons contains what matters. Ernest Sosa and Jeff McMahan provide arguments of this kind.³⁴ Sosa argues that what explains why causal connections are important is that one important causal connection is non-branching survival. Survival for Sosa refers to being the unique closest continuer of a person. McMahan argues that causal connections are important if they have a physical realiser: the continued existence of enough of one's brain. I agree with the spirit of these arguments that something associated with causation explains the causal requirement. However, unlike Sosa and McMahan, I seek to provide an answer that is consistent with Parfit's own view on what matters.

When I introduced quasi-memories and quasi-intentions as examples of first-personal connections, I highlighted that both are defined as causal notions. Quasi-memories and quasi-intentions must both stand in the right kind of causal relation to previous experiences or subsequent actions. Without any causal relations then, there cannot be quasi-memories or quasi-intentions.³⁵ Crucially, in contrast to the generic

³³ Parfit, *Reasons and Persons*, pp. 282-87.

³⁴ Ernest Sosa, "Surviving Matters," *Noûs* 24 (1990): 297-322, at pp. 309-13; and McMahan, *The Ethics of Killing*, pp. 60-66.

³⁵ Alan Sidelle makes a related point about Parfit's discussion of whether what matters is Relation R with any cause, a reliable cause, or its normal cause. Sidelle argues that Parfit's discussion is best understood as rejecting the view that there are any further causal

causal link, it is easy to see why first-personal connections add something significant to mere qualitative similarity. First-personal connections express what distinguishes one's psychology from the psychological make-up of others. We can call this a person's *distinguishing psychology*. Distinguishing psychology is opposed to both *generic psychology* and *core psychology*. Generic psychology refers to the parts of one's psychology that are instantiations of generic psychological features which one shares with others, like character traits or habits. Core psychology refers to the psychological capacities that persons have.³⁶

By conveying one's distinguishing psychology, first-personal connections contain what sets oneself apart from others. They express a non-generic sense of 'you' and demark what is special about you. This links well with what matters. The relation of what matters captures a special bond that we have to persons precisely because of what sets them apart; what makes them different and special. Our distinguishing psychology is thereby closely connected to a sense of self. It provides for the possibility of self-identification. As I discussed earlier, when we self-identify, we acknowledge events or persons in time to be of special importance to us. In the example of mere qualitative similarity, it is this basis for self-identification that is missing. The absence of first-personal connections is the more plausible explanation why the relation with a qualitatively similar person fails to contain what matters.

There is another reason in favour of the view that the absence of first-personal connections satisfactorily explains why the relation of mere qualitative similarity does not contain what matters. When introducing first-personal connections, I highlighted that first-personal connections are connections that are "from the inside" and which provide us with a self-centred perspective on the world. This self-centred perspective is closely related to what matters. It provides us with a perspective from which our projects and ambitions are carried out. The continuation of this perspective provides us with the basis for our special concern with our projects and ambitions.

requirements over and above those inherent in the causal psychological connections that constitute Relation R. See Sidelle, "Parfit on 'the Normal/a Reliable/any Cause of Relation R'".

³⁶ The distinction refines the contrast Unger draws between core psychology and distinctive psychology (Unger, *Identity, Consciousness and Value*, pp. 67-71). Unger somewhat misleadingly uses the term distinctive psychology for both distinguishing and generic psychology.

The first-personal perspective also explains why we are rightly involved and anticipate experience of future person stages. We can anticipate from the first-person perspective.³⁷

We should conclude that first-personal connections are a central component of Relation R. They explain why Relation R requires causal connections between psychological features, provide for a sense of self-identification and provide us with a self-centred scheme from which we experience the world. Given the centrality of first-personal connections for psychological connectedness, we should further conclude that strong connectedness requires at least some first-personal connections. Strong connectedness is in turn needed for psychological continuity. If, following my argument in Section II, psychological continuity is primarily what matters, then non-trivial R-relatedness requires first-personal connections.

We might imagine two persons regularly exchanging quasi-memories, quasi-intentions, and other first-personal connections via telepathy. Similarly, in cases of fission the two resulting persons would share many quasi-memories, quasi-intentions, and other first-personal connections. These two persons would exhibit strong psychological connectedness and continuity. But aside from these science fiction examples, it is hard to see how in our world quasi-memories, or other first-personal connections, could be shared between separate persons. I know of no mechanism in our world that ensures that first-personal memories or intentions can be shared. And I certainly know of no mechanism by which we can share first-personal memories, intentions and so on, over a prolonged period of time. In our world then, strong connectedness, a requirement for continuity, cannot plausibly be met between separate individuals.

IV. Third Argument: Less Importance to Persons

In the previous two sections, I examined and rejected arguments that respectively sought to undermine the unity of the individual and the separateness of persons. We can defend the unity of the individual and the separateness of persons.

³⁷ Cf. Velleman, "Self to Self," pp. 67-76.

I will now examine a third argument. Rather than disputing the unity of the individual or the separateness of persons, it disputes that the separateness of persons or the unity of the individual have moral importance. Parfit's reductionist views on personal identity should give us reason to attach less significance to persons. Parfit argues that a person's existence can be reduced to facts about mental and physical events. Over and above these facts, there does not exist an entity like a Cartesian Ego or a soul.³⁸ Because a person's existence just consists in facts about mental and physical events, there is less that is involved in the fact of personal identity. This should give us grounds to care less about the fact of personal identity.³⁹ The argument relies exclusively on Parfit's reductionist answer to the question of what a person is and does not rely on his more specific claims about what matters for prudential and anticipatory concern. We can still believe that reductionism about persons should lead us to adopt an impersonal morality, even if we think identity is what matters prudentially. This line of argument, while often overlooked, deserves scrutiny.⁴⁰

Let me now turn to the argument. What are the reasons for believing that persons matter less under the reductionist view? Parfit describes facts about personal identity as being a "deeper truth" under the non-reductionist view. He points out that many of us would attach great significance to a separate existence over and above our body and related mental and physical events. Since personal identity is less important, we should also attach less significance to the separateness of our respective existences. This consideration does not seem decisive. Various authors

³⁸ Cf. Parfit, *Reasons and Persons*, pp. 219-28, 236-38, 245-52.

³⁹ Cf. Parfit, *Reasons and Persons*, pp. 337-38, 340-41; and Parfit, "The Unimportance of Identity," pp. 28-41.

⁴⁰ A typical example is David Shoemaker's discussion in the *Stanford Encyclopedia* entry on *Personal Identity and Ethics*. In Section 4 Shoemaker discusses the argument that the adoption of reductionism and the rejection of a deep metaphysical divide between persons could undermine the separateness of persons. But he then writes that the success of such an argument will depend on the kind of psychological or moral unit that the view espouses (cf. David W. Shoemaker, "Personal Identity and Ethics," in *The Stanford Encyclopedia of Philosophy*. Winter 2016 Edition, ed. Edward N. Zalta (URL: <<https://plato.stanford.edu/archives/win2016/entries/identity-ethics/>>)). For a good explanation of the difference between this line of reasoning and other revisionary arguments see Parfit, "Later selves and moral principles," pp. 147-49. For reasons how this type of argument might fill a gap in other revisionary arguments see also John Broome, "Utilitarian Metaphysics," in *Interpersonal Comparisons of Well-Being*, ed. Jon Elster and John E. Roemer (Cambridge: Cambridge University Press, 1991), pp. 70-97.

have pointed out that their belief in the moral (and prudential) importance of persons has not diminished even as they have become convinced of a reductionist picture of the person.⁴¹ Their reason for assigning importance to persons depends on the centrality of persons for our projects and social surroundings. It depends on how persons relate to their future and to others. It never depended on there being a separate entity who is this person.

There is a stronger argument for the reductionist critique. If we are reductionist about personal identity, then we can express every fact about personal identity in another way. We can re-describe these facts as impersonal facts about mental and physical events. But if these facts are just equivalent to the more ordinary, impersonal facts, then it is unclear why we are justified in ascribing greater significance to facts about personal identity than we do to the equivalent impersonal facts. Following this argument, it is not so much the absence of a Cartesian Ego that makes the difference, but rather the availability of an alternative, impersonal description of one's life. If these two ways of describing one's life are indeed equivalent, then we should be suspicious whether the added significance we attach to persons is indeed justified.⁴²

Mark Johnston provides an objection to this kind of argument. He dubs this line of reasoning by Parfit the "argument from below". The argument from below seems to hold that facts about higher level entities are less important as long as they do not involve any superlative, non-reducible entities. It would seem that we can only reason bottom up, from lower level entities and descriptions, and cannot invoke higher level entities in our arguments. The argument from below denies that the value of the whole can be greater than the sum of its parts. If the lower level entities do not carry value, then the higher level entities cannot either. The composition of

⁴¹ Wolf, "Self-Interest and Interest in Selves," esp. pp. 705-8; Adams, "Should Ethics be More Impersonal?," pp. 454-60; and Johnston, "Human Concerns without Superlative Selves," p. 159.

⁴² Parfit makes this argument in *Reasons and Persons* (cf. pp. 340-1), later however Parfit writes that it was misleading to claim that a person's life could be re-described impersonally. Nonetheless, he insists that an impersonal conceptual scheme would be neither scientifically nor metaphysically worse than our current conceptual scheme (Derek Parfit, "Experiences, Subjects, and Conceptual Schemes," *Philosophical Topics* 26 (1999): 217-70).

these entities cannot “create” value. Johnston provides a *reductio* against this argument. Together with physicalism the argument from below implies that the only thing that could matter in our world would be microphysical particles. But evidently these are not, by themselves, of any value. Johnston points out how this is not a proof of moral nihilism but rather a *reductio* against Parfit’s argument from below.⁴³ We can add that under a dualist view, the argument from below would only count mental events or experiences as having importance to us. But very few of us think that the only thing that has value to us are mental states. To make Johnston’s point clearer, we can give examples where Parfit’s reductionist deconstruction seems implausible. We can be reductionist about art and say that the Mona Lisa just consists in a poplar panel and various coloured pigments bound together by oil. Presented this way, it is hard to see why we should attach any significance to the Mona Lisa at all.

Parfit replies to Johnston’s criticism with some examples of his own.⁴⁴ In Parfit’s examples the reductionist strategy seems more plausible. One example is related to the definition of death. Plausibly we are reductionist about death insofar as death just means the cessation of certain functions necessary for our continued existence. According to one view it matters how we define and use the word “death”. But we may plausibly think that what should matter to us morally speaking is which morally important functions cease to exist rather than whether a specific definition is met. Being alive is important only insofar as the functions that constitute “being alive” are important or valuable.⁴⁵ We then need a way to distinguish Parfit’s more plausible examples, like the definition of death, from other examples, like my own about the Mona Lisa. In other words, we need to show that reductionism about death and persons is a different sort of reductionism from the one involved in art.

I have already alluded to one possible answer, the one Parfit wants to defend. In the case of defining death (and personal identity Parfit may add), we are dealing with merely verbal disputes. In the case of art, on the other hand, this is not the case. Parfit writes:

⁴³ Johnston, “Human Concerns without Superlative Selves,” pp. 154-56, 167-68.

⁴⁴ Parfit, “The Unimportance of Identity,” pp. 29-31.

⁴⁵ A more complete defense of this claim is developed by Eric Olson. Eric Olson, “Why Definitions of Death Don’t Matter,” (unpublished manuscript).

“When I claim that personal identity just consists in certain other facts, I have in mind a closer and partly conceptual relation. ... But, if we knew the facts about these [psychological] continuities, and understood the concept of a person, we would thereby know, or would be able to work out, the facts about persons. Hence my claim that, if we know the other facts, questions about personal identity should be taken to be questions, not about reality, but only about our language”.⁴⁶

The most straightforward way to interpret this response is to understand it as analytical reductionism. Analytical reductionism would mean that we could reduce in principle statements involving persons to statements that do not involve persons just in virtue of the meaning of the word “person”. This form of reductionism seems plausible in Parfit’s cases that concern the definitions of words. Analytical reductionism would, however, also mean that the statement about persons and the impersonal statement to which it can be reduced necessarily have the same truth-value. If the difference is merely about our language and not about reality, then the relation of equivalence between a statement about persons and an impersonal statement should hold necessarily. But here Parfit provides the best counterexamples against himself. Reductionism about persons does not hold necessarily, non-reductionism may well have been true. If we had evidence of persons remembering events from distant times and were these events confirmed, this would support the case for an immortal soul that can be reincarnated.⁴⁷ Reductionism does then not hold as an analytical necessity.

Rather than analytical reductionism, Parfit ought to hold that reductionism about persons is ontological reductionism. Here the idea is that we can translate facts about a specific entity into facts that do not presuppose this entity. Instead of persons we can talk of mental and physical events and their relations. Instead of the Mona

⁴⁶ Parfit, “The Unimportance of Identity,” p. 33. Elsewhere Parfit writes that under his view the existence of persons is only “a fact of grammar” (“Later selves and moral principles,” p. 158), he also writes that most facts about persons only exist because of *the way we talk* (*Reasons and Persons*, pp. 223, 226, 341). Parfit also defends more explicitly the view that an impersonal conceptual scheme would be no worse than our current conceptual scheme (“Experiences, Subjects, and Conceptual Schemes”).

⁴⁷ Cf. Parfit, *Reasons and Persons*, pp. 227-28.

Lisa we can talk of colour pigments and their spatial relations, and so on. One specific form of ontological reductionism is constitutive reductionism. Under constitutive reductionism some entities constitute others. A common example for constitutive reductionism are clay statues. The statue does not exist independently from the lump of clay, but neither is it identical to it. Rather the lump of clay constitutes the statue. In cases of constitutive reductionism, we would still say that there is an additional entity in the world. The statue does exist in the world and has an existence separate but not independent from the lump of clay. The existence of the statue will always be parasitic on the existence of the clay. But we can destroy the statue without destroying the lump of clay. This gives us a strong sense how the statue is a separate entity. Facts that hold about the statue are therefore not merely conceptual facts about how to use words, they are facts about a really existing entity. Parfit claims to be a realist about importance by which he means that he attaches importance only to those facts that are ontologically real. But if Parfit is a realist about importance in this way, then he should attach significance to constituted entities. Constituted entities are ontologically real after all. Given that he does not attach significance to persons, his reductionism is most plausibly not a constitutive one.⁴⁸

Instead of constitutive reductionism, Parfit needs to invoke eliminativist reductionism. According to eliminativist reductionism, the reduced entity does not really exist. It is not part of one's privileged ontology. Instead, we only have terms of convenience that do not refer to any real entity at all. This interpretation gives a strong sense that we would be treating language as more important than reality if we attached significance to persons. The problem with this reading is that Parfit does not give any argument for eliminative reductionism about persons. His reductionist

⁴⁸ Here things are getting confused since Parfit does expressly claim to be a constitutive reductionist about persons ("The Unimportance of Identity," pp. 16-17; and "Experiences, Subjects, and Conceptual Schemes," p. 218). However, he describes facts about persons as merely conceptual facts. It might be that I have overlooked something in my argument and that some forms of constitutive reductionism give rise to genuine entities with facts about reality (like statues, art works and so on) while other forms of constitutive reductionism give rise to merely conceptual facts. Parfit in any case fails to make this argument and I do not know of any good argument to this effect. See also David Shoemaker's post and ensuing discussion on the PEA Soup blog for more detail on this discussion. David W. Shoemaker, "Parfit's 'Argument from Below' vs. Johnston's 'Argument from Above'," in *PEA Soup Blog* (2006, URL: <https://peasoup.typepad.com/peasoup/2006/04/parfits_argumen.html>).

arguments seek to establish that it is possible to give an impersonal description of one's life and that no appeal to a higher entity is needed. But, of course, the ability to use a different vocabulary does not establish the need to use it. We can similarly give a description of an artwork without mentioning its existence, but this does not mean that we should not include the artwork in our ontology.

We need a different way to distinguish between cases where reductionism does disenchant our ordinary concepts and those cases where it does not. One possible explanation is that in some cases the relations between constituent entities have significance over and above the entities while in other cases they do not. In the example of reductionism about art, it is the special way how the colour pigments of the Mona Lisa stand to one another that makes the Mona Lisa important over and above its individual elements. If we could show that the relations of individual events are not significant in the case of persons, then we would have achieved a reductionist debunking of our concept of a person.

John Broome provides such an argument.⁴⁹ Broome wants to argue that the relations between the different stages of a person are axiologically insignificant. There would be just as much value in the world regardless of the relation between person stages. Broome draws a comparison between a world with one person and a world with two persons that correspond to the two halves of the first person's life. We can imagine that the two persons are living two different causally isolated lives that correspond to all of the person stages that form part of the first half or the second half of the first person's life respectively. In this situation Broome says that it is unclear why the world with one person is any different in terms of value from the world with two persons.

Broome's argument should equally hold if we decompose the person's life further into different time-slices rather than comparably big units. If we decompose a life to this extent however it is difficult to see how the importance of fulfilling desires or achieving projects can be captured. Time-slice persons are not extended beyond an ephemeral moment. The satisfaction of desires and the accomplishment of projects however extends in time. Parfit's proposal of the success theory of well-

⁴⁹ Broome, "Utilitarian Metaphysics," pp. 87-90.

being makes this particularly clear. A success requires some extension in time and does not refer merely to someone being a specific way at a given time.⁵⁰ On most accounts of a person's well-being, we consider projects or desires to be an important component. Since projects and desires require continued existence over at least some time, the relation between the individual constitutive parts of a person's life does have axiological significance, contrary to what Broome argues.

We should conclude that Broome's argument fails as well. Neither Parfit's claims about "less deep" truths of personal identity, nor Parfit's appeal to "merely conceptual truth", nor Broome's argument about the axiological insignificance of the relations that unite a life have succeeded. None of the three arguments has established sufficient ground to reject the importance of persons based on a reductionist metaphysics of persons.

V. Conclusion

I have argued that Parfit's step from the questions of personal identity and of what matters for self-interest to the question of what matters for our moral theorizing is not warranted. We can grant Parfit's answer to the question of what matters without having to adjust our moral theories. There is no need to engage in the complex discussion over whether or not identity does or does not matter, if we simply want to defend a person-based form of morality. I have defended the unity of the individual and the separateness of persons against Parfit's challenge. Once we see that psychological continuity and not connectedness is the primary part of Relation R, his challenge fails. And once we understand that Parfit's reductionism about persons is best understood as a constitutive reductionism, we realize that the unity of the individual and separateness of persons have the same significance as they had before Parfit's challenge.

Parfit's contributions to the metaphysics of personal identity and its implications for rationality and self-interest are truly outstanding. For a while I feared that Parfit provided a strong challenge to my moral beliefs as well. The arguments in

⁵⁰ Parfit, *Reasons and Persons*, pp. 494-99.

this chapter have convinced me otherwise. It seems to me that Parfit's argument for a revisionary understanding of morality fails. I can be reassured. It does not matter what matters.

Chapter 2. Separate Persons Behind the Veil

I. Introduction

The *separateness of persons* and the *veil of ignorance* are two among many arguments or devices that were made prominent by John Rawls in *A Theory of Justice*. Neither of the two originated with Rawls and versions of them existed prior to Rawls.¹ But Rawls's formulation and the integration of these ideas in a comprehensive theory of justice made them prominent and canonical.

I want to explore the relation between the separateness of persons and the veil of ignorance. In particular, I want to assess whether the separateness of persons gives us a reason *not* to use the veil of ignorance. In doing so, I will draw on the distinction that I outlined in the introduction to this dissertation between a justificatory requirement and a substantive requirement. I will assess different constructions of the veil of ignorance, John Harsanyi's, Rawls's own, and Ronald Dworkin's, in light of the requirements of the separateness of persons. The discussion of these examples will help me to argue how the veil of ignorance can meet the two requirements that the separateness of persons imposes. Yet this requires that the veil both plays a different role and is constructed differently from the standard model that Rawls made popular. I will first discuss the justificatory requirement (Section II) before discussing the substantive requirement (Section III).

¹ I have traced origins of the separateness of persons objection in the Introduction. The veil of ignorance is used before by John Harsanyi. ("Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy* 61 (1953): 434-35; and "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63 (1955): 309-21.) The first use of the veil of ignorance precedes even Harsanyi's widely known use. William Vickrey briefly referred to the idea that a distribution of income should be distributed in accordance with the choice an individual would make not knowing his position in society ("Measuring Marginal Utility by Reactions to Risk," *Econometrica* 13 (1945): 319-33, at pp. 328-29.)

II. Justification for Separate Persons

A. *The Impartial Spectator*

I will begin by discussing how the veil of ignorance features in the justification of principles of justice. A good starting point for this is the impartial spectator. Rawls reconstructs an argument in favor of utilitarianism as a principle of justice that makes use of the impartial spectator. He considers this the “most natural way” to argue for utilitarianism.² The argument shares a common core with Rawls’s own contractualism insofar as utilitarianism is justified by appeal to a principle of justifiability. This contrasts with a teleological justification for utilitarianism which simply appeals to the goodness of welfare directly. The standard for justifiability proposed is the endorsement by an ideally rational and impartial spectator. This method is inspired by David Hume’s and Adam Smith’s use of the “judicious” (in Hume’s case) or “impartial” (in Smith’s case) spectator, even though Rawls notes that Hume’s own moral philosophy is not utilitarian.³

The argument for utilitarianism is then the following. The impartial spectator imagines herself to be in each person’s position. She is equally sympathetic to everyone’s plight. She imagines to be in each person’s position and what it is like to be in this position. After imagining herself to be in each person’s position, she approves or disapproves of principles of justice. Approval is understood here as a kind of pleasure that arises from the reproduction of the experience of each person. The spectator would therefore feel the experience of each person with positive experiences and negative experiences canceling each other out. In the end the impartial spectator would endorse the principle of justice that brings about the largest

² Rawls, *A Theory of Justice*, rev. edn., pp. 23-24.

³ David Hume, *A Treatise of Human Nature*, ed. L.A. Selby-Bigge (Oxford: Clarendon Press, 1888), pp. 574-91; Adam Smith, “The Theory of Moral Sentiments,” in *The Essential Adam Smith*, ed. Robert L. Heilbroner (Oxford: Oxford University Press, 1986), pp. 57-147, at pp. 93-96, 100-9, 118-123. For Rawls’s discussion see Rawls, *A Theory of Justice*, rev. edn., pp. 28-29; and John Rawls, *Lectures on the History of Moral Philosophy* (Cambridge, MA.: Harvard University Press, 2000), pp. 84-100. He does not discuss Smith’s moral theory in any depth.

net sum total of welfare.⁴ But this argument does not take notice of the separateness of persons. The impartial spectator, by balancing out the different experiences of different persons, treats them as if they formed part of one system of experiences or desires. The sympathetic imagination of the impartial spectator treats all of these lives as stages of a very long life of one spectator.

The problem is that this interpretation of the method of the impartial spectator achieves impartiality only by subsuming the various individual points of view of separate individuals under one overarching point of view. Rawls does not want to sacrifice impartiality and he is right to insist on it. While partiality has an important role in personal morality, its role in matters of justice is dubious. As a principle of justice rather than as a moral theory, utilitarianism is relatively more plausible. The objection that utilitarianism cannot allow for the personal perspective by ruling out partiality does not apply when we restrict utilitarianism to be a principle of distributive justice alone. While individuals do not have to treat everyone with equal concern and respect, governments have to. Equal concern is, in Ronald Dworkin's famous words, the sovereign virtue of political community.⁵ While impartiality has to be observed in distributive justice, impartiality is not the same as the impersonality or the monopersonality of an impartial spectator.

B. *John Harsanyi*

John Harsanyi provides us with a second example of a veil of ignorance argument. Like the impartial spectator, it is aimed to support a version of utilitarianism. As in the case of the impartial spectator, Harsanyi makes use of the veil of ignorance because he wants to find impartial principles of distributive justice. While ordinary judgments about distributive justice are clouded by our particular situation, truly impartial judgments will be found in the absence of knowledge of our particular situation. Therefore, he imagines the hypothetical choice an impartial

⁴ Rawls, *A Theory of Justice*, rev. edn., pp. 161-65.

⁵ See the title and the introduction to *Sovereign Virtue* (Cambridge, MA.: Harvard University Press, 2000).

observer would make not knowing his or her position in society. She would assign an equal probability to being in each person's position and then choose accordingly. A crucial difference to the impartial spectator is that Harsanyi does not imagine his impartial observer to be guided by a hedonic evaluation of each person's life. Instead, Harsanyi imagines the impartial observer to occupy each person's position *with their conception of the good and their preferences*. Harsanyi then reasons that if the impartial observer follows expected utility theory, then she would systematically choose the distribution that maximizes average utility.⁶

What does it mean, however, to be in a person's position with their conception of the good? A person's position includes all social facts that are potentially morally relevant and a person's conception of the good which is deeply tied up with their psychology and personality. An apparent interpretation of being in a person's position with their conception of the good is, therefore, to *be* this person.⁷ Yet this interpretation is puzzling. It is impossible for the impartial observer to be numerically identical with another person.

Perhaps what Harsanyi's veil of ignorance argument requires is not numerical identity but something similar to it. One proposal is what Bernard Williams calls participation imagery.⁸ Williams discusses the case of a person imagining that he is Napoleon, for example in a dream. This person would have imagery from Napoleon's first-person point of view. Such imagery is different from visualizing Napoleon at the battle of Austerlitz. The imagery is "from the inside", it occupies Napoleon's point of view. Such imagery is clearly coherent and possible. It occurs when we dream, role-play or enact dramatic performances. Participation imagery can give us, furthermore, a good sense of someone's hedonic state. The impartial spectator that I discussed earlier, for example, can make use of participation imagery in order to replicate the

⁶ Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking"; and Harsanyi, "Morality and the theory of rational behavior," in Sen and Williams, *Utilitarianism and beyond*, pp. 631-38.

⁷ See Matthew D. Adler, *Well-Being and Fair Distribution* (Oxford: Oxford University Press, 2012), pp. 198-99; and Hilary Greaves and Harvey Lederman, "Extended Preferences and Interpersonal Comparisons of Well-Being," *Philosophy and Phenomenological Research* 96 (2018): 636-67, at pp. 640-42.

⁸ Williams, *Problems of the Self*, ch. 3.

first-personal experience of different persons. Harsanyi's impartial observer is not guided by a hedonic evaluation, however. This means that participation imagery is not enough. Having a certain set of experiences and beliefs from a first-personal point of view allows us only to rank alternatives within one person's life. But the impartial spectator would not be able to compare alternatives across lives. This way the impartial spectator could not choose a principle of justice. If one principle is better for A and worse for B and another principle is better for B and worse for A, the impartial spectator cannot tell which principle would be better in expectation.

A second proposal is to make use of the phenomenon of *de se* beliefs. Take the case of mad Heimson who believes he is Hume, yet who fails to believe the proposition that "Heimson is Hume". Heimson believes *de se* that he is Hume yet believes *de dicto* that "Heimson" is not Hume. A *de se* belief is a belief that belongs to a self-centered scheme of reference, a first-personal point of view. Perhaps this means that we can have *de se* and *de dicto* preferences as well. Greaves and Lederman argue that such *de se* preferences do not require that subjects are deluded about their identity as Heimson is.⁹ Their argument relies on the idea that an agent can have a preference to be cured from an illness even when she knows that she will remain sick. Preferences, they argue, can be held over alternatives that do not obtain. However, this argument sidesteps the important challenge. While we can have preferences over possibilities we know that do not obtain, invoking *de se* attitudes does nothing to alleviate the concern that one cannot have preferences over scenarios one *knows to be impossible*.

The argument reveals an important possible avenue. Perhaps we should take the label "veil of ignorance" more seriously. According to this interpretation, the thought experiment requires a straightforward deprivation of knowledge about oneself. We can imagine it as a form of transient amnesia. The impartial observer has temporarily forgotten who she is. Perhaps she consumes a new neurological drug which temporarily blocks the access to her psychological make-up, except for her rational reasoning capacities. She can only assign an equal probability to being any

⁹ Greaves and Lederman, "Extended Preferences and Interpersonal Comparisons of Well-Being," pp. 644-50.

given individual in society with their social position and conception of the good.¹⁰ A question for this interpretation is whether it is compatible with all plausible views on personal identity. In particular, there is a concern that psychological, or Lockean, views on personal identity struggle to accommodate the choice of the impartial observer. I discussed Parfit's version of a Lockean view in the previous chapter. The concern is that there is a complete breakdown of psychological connectedness between the impartial observer behind the veil of ignorance and the impartial observer once the veil of ignorance is lifted. If Parfit is correct that Relation R is what matters for prudential concern, then the impartial observer behind the veil of ignorance may have no concern for any of the possible persons she may turn out to be.¹¹

I believe that this concern is unfounded. What the thought experiment requires is that the impartial observer has short-term, transient amnesia. The observer will recover all of her psychological features once the veil is lifted. The objection to Harsanyi's veil stands only if no Lockean view on personal identity can account for transient amnesia. However, Lockean views should be able to say something about transient amnesia and I believe they can. Cognitive psychologists distinguish between three components of our faculties of memories. One is the *encoding* of new memories, a second is the *storage* of memories, a third is the *retrieval* of memories. In cases of transient amnesia, the storage of memories is unaffected and only the retrieval is temporarily blocked. In order to account for the numerical identity of individuals pre- and post-amnesia, a Lockean can argue that continuity of storage of memories is sufficient.¹² A Lockean need not even argue that continuity of storage is always sufficient, but only in cases where the retrieval of information is blocked for a short period. A similar revision would seem necessary if Lockeans want to account

¹⁰ Alex Voorhoeve, "Matthew D. Adler: Well-being and fair distribution: beyond cost-benefit analysis," *Social Choice and Welfare* 42 (2014): 245-54, at pp. 247-48. Voorhoeve credits Michael Otsuka for the thought experiment.

¹¹ For the concern see Voorhoeve, "Matthew D. Adler: Well-being and fair distribution," p. 248.

¹² See also Andreas L. Mogensen, "The Brave Officer Rides Again," *Erkenntnis* 83 (2018): 315-29, at pp. 318-19; Jamies Baillie, "Recent Work on Personal Identity," *Philosophical Books* 4 (1993): 193-206, at p. 195.

for cases where persons are put in a temporary coma and upon reawaking show a perfect psychological connection, or even the more mundane case of a person sleeping and having no active psychological connections for the duration of the sleep.¹³

While Harsanyi's thought experiment so construed is coherent, it clearly violates the justificatory version of the separateness of persons. The impartial observer is very similar in this regard to the impartial spectator. There are three important differences. The impartial observer is rationally self-interested rather than sympathetic. Second, the observer is also not perfectly knowledgeable since she lacks knowledge of who she is and is assumed to have an equal probability of being each person. Third, while the impartial spectator imagines herself to be in each person's position in turn, the impartial observer considers each person's life to be a possible future life of herself. But notwithstanding these differences, Harsanyi's first argument falls foul of the separateness of persons. It merges all individuals together by considering them as mere possibilities of one person's future.

While the impartial observer form of the argument is prominent, there is another interpretation of Harsanyi's point.¹⁴ Harsanyi repeatedly speaks of "ethical preferences" when talking about the judgment that a society with higher average

¹³ Andrew Brennan thinks that we should therefore switch to a more coarse-grained understanding of psychological connections in which episodes separated by more than moments are connected in time. Andrew Brennan, "Amnesia and Psychological Continuity," *Canadian Journal of Philosophy* 15 (1985): 195-209, at pp. 196-97.

¹⁴ For the above, impartial observer interpretation see Philippe Mongin, "The Impartial Observer Theorem of Social Ethics," *Economics & Philosophy* 17 (2001): 147-79; John E. Roemer, "Harsanyi's Impartial Observer is not a Utilitarian," in *Justice, Political Liberalism, and Utilitarianism*, ed. Marc Fleurbaey, Maurice Salles, and John Weymark (Cambridge: Cambridge University Press, 2008), pp. 129-35; Marc Fleurbaey, "Economics and Economic Justice," in *The Stanford Encyclopedia of Philosophy*. Winter 2016 Edition, ed. Edward N. Zalta (URL: <<https://plato.stanford.edu/archives/win2016/entries/economic-justice/>>), sec. 3; and Hilary Greaves, "A Reconsideration of the Harsanyi-Sen-Weymark Debate on Utilitarianism," *Utilitas* 29 (2017): 175-213, at pp. 179-81. Harsanyi uses the language of the impartial observer in later statements of his argument ("Morality and the theory of rational behavior" (in 1977)) but not in earlier ones. For the second, ethical preferences interpretation see Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking"; Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," pp. 314-16; John C. Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory," *American Political Science Review* 69 (1975): 594-606, at p. 598; and Lara Buchak, "Taking Risks behind the Veil of Ignorance," *Ethics* 127 (2017): 610-44, at pp. 633-35.

utility is superior to one with lower. Impartiality is an ingredient of ethical preferences and the choice behind the veil of ignorance brings out this aspect of ethical preferences. Harsanyi gives the following example to illustrate this. A wealthy capitalist may prefer capitalism over socialism because she is better off under capitalism than under socialism. If, however, the person would prefer capitalism over socialism regardless of her social position, then this indicates an ethical stance in favor of capitalism.¹⁵ The difference to the previous interpretation is that Harsanyi makes a claim about the considered ethical judgments of everyone, not about the judgment of one impartial observer. It is this second interpretation which can be identified with a form of contractualism. The ethical preference interpretation of Harsanyi's veil of ignorance gives a foundational role to agreement between different persons. Principles of justice are true because they would be agreed upon behind a veil of ignorance which brings out our ethical preferences. The impartial observer interpretation does not give importance to agreement. Instead, it is the endorsement by an impartial observer which gives validity to principles of justice.

Harsanyi's second argument therefore faces the difficulty of explaining why the parties behind the veil of ignorance would achieve unanimous agreement. The answer is that the veil of ignorance brings to the forefront individuals' extended preferences.¹⁶ An extended preference is a meta-preference concerning a pair of social conditions and ordinary preferences. For example, preferring x (being a monk, having a religious conception of the good) over y (being a surfer, having an Epicurean conception of the good) is an extended preference. If everyone's extended preferences are the same, then we can explain why there would be unanimous agreement behind the veil of ignorance. There would be only one shared extended preference function to be taken into account.

¹⁵ Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality?," p. 598. See also Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking"; Harsanyi, *Morality and the theory of rational behavior*, pp. 631-32.

¹⁶ For extended preferences see Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," pp. 316-19; and John C. Harsanyi, *Rational behavior and bargaining equilibrium in games and social situations* (Cambridge: Cambridge University Press, 1977), pp. 51-60.

Harsanyi justifies the assumption of identical extended preferences with the following argument.¹⁷ If we observe a difference in preferences between two individuals, then there will be some cause for this divergence. If we identify the cause and make the cause part of the object of preferences, then we eliminate this difference. We proceed until all differences in preference are accounted for. In some cases, this causal argument is convincing. The fact that one person ranks living as a monk over living as a surfer can plausibly be explained by the fact that this person has a religious conception of the good. The causal conditions for preferring x (being a monk, having a religious conception of the good) over y (being a surfer, having an Epicurean conception of the good) are more difficult to account for. Once we incorporate all relevant causes for our conceptions of the good, it is difficult to imagine that there is much left of a personality or agency.

This shows that the assumption that there is only one shared extended preference function denies the individuality and separateness of persons.¹⁸ The parties choosing behind the veil of ignorance become indistinguishable. The thought experiment reduces individuals to an abstract preference relation without any individuality. By claiming that everyone shares the same extended preference relation, it furthermore denies the plurality of conceptions of the good. The extended preference relation encapsulates a form of second-order preferences. Such second-order preferences are an important part of one's conception of the good. They determine which parts of our first-order pursuits of the good we reflectively endorse and which first-order pursuits of the good we want to rid ourselves of.¹⁹ Furthermore, second-order preferences determine whether we attach value to living lives authentically in accordance with our preferences or whether we believe that a life is

¹⁷ Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," pp. 316-19. A clear version of this argument is made by Serge-Christophe Kolm, *Justice and Equity* (Cambridge, MA.: The MIT Press, 1997), pp. 165-67. It is also hinted at by Jan Tinbergen, "Welfare Economics and Income Distribution," *American Economic Review* 47 (1957): 490-503, at pp. 500-1.

¹⁸ See John Rawls, "Social unity and primary goods," in Sen and Williams, *Utilitarianism and beyond*, pp. 173-83.

¹⁹ See e.g. Harry Frankfurt, "The Importance of What We Care About," *Synthese* 53 (1982): 257-72 in which Frankfurt draws a contrast between caring which is intimately connected to our idea of a good life and mere liking or wanting.

lived well if it has objective goods even if these are not valued. Two individuals may share the preference of x (being a monk, having a religious conception of the good) over y (being a surfer, having an Epicurean conception of the good), while differing about the choice between x^* (being a monk, having an Epicurean conception of the good) over y (being a surfer, having an Epicurean conception of the good). On one second-order conception of the good, a life of religious enlightenment is always better than a life of hedonic pursuits. On another second-order conception of the good, while religious enlightenment is objectively good, it has to be pursued authentically.

Harsanyi's ethical preference argument thereby denies the separateness of persons. The impartial observer does so by turning individuals into a possible future of a single decision-maker. Ethical preferences do so by stripping individuals of all their individuality and distinctiveness. *Every* person is now overlooking the boundaries between persons by considering every person's life to be a mere possibility of one's future.

C. John Rawls

Rawls's use of the veil of ignorance is motivated by the desire to respect the separateness of persons. The impartial spectator failed to do so. It assumed perfect knowledge, perfect sympathy, and perfect imaginative powers. These conditions lead to a conception of impartiality that identifies impartiality with impersonality. Impersonality in turn means the conflation of all desires into a system of desires assessed by the impartial spectator. This conflation in turn violates the separateness of persons.²⁰ Rawls wants to develop an alternative to the impartial spectator in the social contract tradition. He replaces perfect sympathy with mutual disinterest and rational self-interest. This requires a relaxation of the condition of perfect knowledge given that Rawls wants to retain impartiality. The parties in the original position should not be able to rig principles in their favor. The solution is the veil of ignorance which deprives individuals of all knowledge that may allow them to tailor principles of justice in their favor.

²⁰ Rawls, *A Theory of Justice*, rev. edn., pp. 164-66.

There are three important differences to Harsanyi's veil of ignorance and how it is employed. First, the two methods are concerned with different objects of choice. Harsanyi is looking for a broad moral principle like utilitarianism. Rawls is looking for principles of justice that regulate the basic structure of society.²¹ Second, the grounds for choice are different. The parties behind Harsanyi's veil are choosing based on welfare considerations. The parties behind Rawls's veil are deprived of the knowledge of their particular conception of the good. Unlike in Harsanyi's model where the parties choose according to a higher-order conception of the good, in Rawls's model the only goods that can influence their choice will be goods they know will be important to them regardless of their conception of the good; i.e. primary goods. Third, Harsanyi assumes that every person has an equal probability of being in each position. Rawls, on the other side, deprives the parties of the original position of any knowledge of probabilities.²²

Rawls's veil of ignorance ensures that the parties in the original position will agree on principles of justice. The veil of ignorance deprives them of all information that would allow any differentiation between the parties. Since the parties are also equally rational, they will all choose the same. Rawls even writes that "[therefore], we can view the choice in the original position from the standpoint of one person

²¹ This difference becomes apparent in an exchange between Harsanyi and Rawls in which Harsanyi seems to assume that Rawls is advocating for maximin as a general distributive principle as opposed to a principle for the regulation of the basic structure of society. John Rawls, "Some Reasons for the Maximin Criterion," *American Economic Review* 64 (1974): 141-46; and Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality?," pp. 605-6.

²² The third difference has raised a lot of controversy. Critics have suggested that it violates accepted standards of rationality for the parties in the original position not to make the equiprobability assumption. (E.g. Kenneth J. Arrow, "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice," *Journal of Philosophy* 70 (1973): 245-63, at pp. 249-52; Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality?," pp. 598-600.) Rawls's departure from Harsanyi should however be seen in light of his design of the original position. The original position is set up in a way that achieves a reflective equilibrium between the theoretical construction and our intuitions about justice. Rawls gives therefore many moral arguments for the design of the original position. He appeals to the strains of commitment, stability, and self-respect. All these moral arguments strengthen the design of the original position in a way that excludes the assumption of equal probabilities. In other words, the debate around the equal probability assumption should not be narrowed to a debate about rational choice theory. See also Samuel Scheffler, "Rawls and Utilitarianism," in *The Cambridge Companion to Rawls*, ed. Samuel Freeman (Cambridge: Cambridge University Press, 2003), pp. 426-59, at pp. 433-36; and Michael Moehler, "The Rawls-Harsanyi Dispute: A Moral Point of View," *Pacific Philosophical Quarterly* 99 (2018): 82-99.

selected at random".²³ Rawls goes on to explain that this step is needed if we are to insist on unanimous agreement in the original position. Without ensuring that all parties will choose the same, we would not be able to work out a theory of justice at all. The problem in the original position would become hopelessly complicated.²⁴

These remarks indicate an important commonality with Harsanyi's veil of ignorance. In both cases, the parties are deprived of information to an extent which makes it impossible to differentiate between them. For the purposes of choice, we could simply use a single person. In Harsanyi's argument this becomes implicit in the impartial observer interpretation. In Rawls's argument this is implicit. For this single person, furthermore, the choice becomes one where the person does not know which position in society she will inhabit. In effect, all positions in society are transformed into possibilities from the decision-maker's point of view. Rawls is then guilty of the same charge as Harsanyi. His veil of ignorance violates the justificatory version of the separateness of persons. All lives are seen as merely possible futures from the point of view of the decision-maker.²⁵

The diagnosis of this mistake is important. Rawls tries to develop a procedure by which principles of justice can be evaluated. He is motivated not to repeat the mistake of the impartial spectator. His proposal therefore does not imagine that all lives are lived by the impartial spectator in seriatim. Unlike Harsanyi's impartial observer interpretation, he furthermore does not imagine one impartial observer who is equally likely to be each member of society. For Rawls, as for Harsanyi's ethical preference interpretation, it counts that each member of the society would hypothetically consent to the principles of justice. The problem he then faces is how to ensure unanimity. In order to solve this problem Rawls requires a veil of ignorance which deprives individuals of all information that could create disagreement. But in doing so, he also deprives the members of society of their separateness from one another.

²³ John Rawls, *A Theory of Justice* (Cambridge, MA.: Harvard University Press, 1971), p. 139. Rawls changes the word "choice" to "agreement" in the revised edition (p. 120).

²⁴ Rawls, *A Theory of Justice*, rev. edn., pp. 121-23.

²⁵ See also Nagel, *The Possibility of Altruism*, pp. 138-40.

D. From Rawls to Dworkin: The Nature of the Veil

The main problem of Rawls's veil of ignorance is that it deprives the parties of the original position of all the knowledge that would allow us to differentiate between them. Ronald Dworkin avoids this problem with his veil of ignorance. Dworkin's veil of ignorance is integrated in his theory of equality of resources. The starting point for equality of resources is the envy test. According to the envy test a distribution of impersonal resources is equal if no one would prefer to trade her bundle of resources with anyone else's bundle of resources. Equality of resources is, however, not exhausted by the envy test. We could fulfil the envy test by giving everyone a certain amount of a resource no one wants. While everyone would be equally miserable this would not be a distribution that shows equal concern, or indeed any concern, for everyone. Dworkin therefore supplements the envy test with an initial auction of resources. He imagines that every member of society has an equal opportunity to auction resources and to later trade them. After all trades have taken place the distribution fulfills the envy test and reflects the tastes of the members of society.

This general ideal has to be revised, however. Individuals in the real world do not start with equal shares and their shares may later decrease through no fault of their own. In order to accommodate these instances of bad brute luck, Dworkin devises a hypothetical insurance scheme. He asks how people would insure themselves against bad brute luck in just circumstances. This is the stage where the veil of ignorance comes into play. The question must be what individuals would be willing to pay for insurance if they did not know their special risk.²⁶

In setting up his veil of ignorance Dworkin departs in one crucial aspect from Rawls. Dworkin does not include a person's talents as being among the information concealed by the veil of ignorance. Talents, he argues, are too closely connected to a person's personality. Without this basis of a person's personality we cannot judge the

²⁶ See Dworkin, *Sovereign Virtue*, pp. 65-79. Dworkin develops this system further into a model for a tax system that mirrors insurance people would take out against their talents yielding fewer resources (Dworkin, *Sovereign Virtue*, pp. 85-109).

ambitions an individual may have. Equality of resources aims at respecting people's responsibility for their shares of resources. Different ambitions should therefore be reflected in the final shares of resources and information about people's ambitions will be crucial.

It is unclear, however, if this reason sufficiently motivates the inclusion of knowledge of talents. It would seem to be possible to construct a veil of ignorance that allows for some knowledge about one's ambitions without allowing for information about talents. In order for a theory to be ambition-sensitive, it must allow for knowledge of a person's ambitions. But it is not required that the person knows her own personality. Only the part of her personality connected to her ambitions must be known to her. Dworkin's insistence on allowing knowledge of one's personality makes more sense, however, as a response to respecting the separateness of persons. It responds to the problem of Rawls's veil where individuals become indistinguishable behind the veil. For the purpose of the choice behind the veil, their separateness does not matter. Dworkin's veil on the other hand, in allowing for the knowledge of one's personality, does not make this mistake. The knowledge of one's personality gives a robust guarantee to the separateness of persons. It avoids the effect that we can see the decision in the hypothetical insurance market simply from the point of one representative individual.

E. From Harsanyi and Rawls to Dworkin: The Justificatory Role of the Veil

I have criticized Harsanyi and Rawls for violating the separateness of persons. This might be surprising, not only because Rawls himself made the use of the separateness of persons objection prominent. It might also be surprising because both philosophers are contractualists. Harsanyi's justification, for example, has been held to be compliant with the separateness of persons for this reason. Contractualism, by its focus on the justifiability to each, embodies a model of justification which gives an equal and separate voice to everyone. Since everyone has a veto, the separateness of persons is respected. It appears that contractualism complies with the individualist

restriction that insists that each person's claims will be counted separately and only include their interests and claims.²⁷

The general idea of this resurrection of Harsanyi is correct. A method of justification that requires justifiability to each would comply with the separateness of persons. But this does not mean that every form of contractualism meets this standard. In Harsanyi's ethical preferences different lives are not recognized as such but transformed into possible lives of the decision-maker. The same happens with Rawls's argument. Since the parties behind the veil become indistinguishable, what counts is the individual calculation of a single member of society.

We can understand this failure in terms of two different models of unanimity rules that are introduced by Thomas Nagel.²⁸ One way to ensure unanimity is to prescribe one course of reasoning for everyone. This is, in effect what Harsanyi and Rawls are doing. But there is another model of unanimity rules. In this model different persons converge from different starting points by modulating their claims and expectations. Such a reconciliation would have to happen without a veil of ignorance. Nagel expresses skepticism about convergence within a model of agents pursuing rational self-interest.²⁹ Without a veil of ignorance the model would struggle to explain why we ought to reconcile our claims with those whom we could oppress. If the motivation is purely the pursuit of rational self-interest, then this might be the most convenient solution for us. Instead, Nagel points to a solution that does not rely on rational self-interest.

²⁷ Hirose, *Moral Aggregation*, pp. 78-84. Hirose does not endorse this argument. Instead, he rather suggests various ways in which one may think the separateness of persons has to be respected. His ultimate conclusion is that none of these ways bar interpersonal aggregation. For this conclusion, Hirose does not need to endorse any particular separateness of persons argument. For the individualist restriction see Derek Parfit, *On What Matters*, vol. 2 (Oxford: Oxford University Press, 2011), pp. 193-96 and Michael Otsuka, "Saving Lives, Moral Theory, and the Claims of Individuals," *Philosophy & Public Affairs* 34 (2006): 109-35, at p. 125. In Footnote 27 Otsuka raises an objection to this idea that is similar to the one I raise in the following paragraph.

²⁸ Nagel, *Equality and Partiality*, pp. 33-40.

²⁹ He mentions David Gauthier's approach of using bargaining theory to determine a contractualist moral theory. See David Gauthier, *Morals by Agreement* (Oxford: Oxford University Press, 1986).

Nagel's solution points at a form of contractualism that is associated with T.M. Scanlon. Scanlon's contractualism is concerned with what it is reasonable to accept (or not reject) as opposed to what it is rational (i.e. in one's self-interest) to accept.³⁰ Scanlon's criticism of Rawls's form of contractualism gives us a sense how the veil of ignorance can still play a role in such a theory.³¹ Scanlon correctly identifies that Rawls's main motivation behind the veil of ignorance is to ensure impartiality. A principle that is impartially acceptable is a principle that can be accepted from every possible standpoint. Therefore, it is nothing about a particular position, a particular conception of the good, or the like, that makes the principle acceptable. The veil of ignorance is quite helpful as such a thought experiment. When we abstract from our peculiarities, we can see whether or not the principle is still acceptable to us. If a principle meets the test of the veil of ignorance, it is because everyone has reason to accept this principle. This argument, however, is very different from Rawls's argument of self-interested choice behind the veil of ignorance. Scanlon's reconstruction of the impartiality argument makes no appeal to self-interest. Instead, it focuses on the reasons individuals have for accepting (or not rejecting) principles. Arguments in Scanlonian contractualism do not admit to a simple reduction of moral questions to prudential questions.

This criticism goes farther than the point previously considered. It holds not only that the veil of ignorance must ensure that the parties behind it are distinct individuals. It also criticizes that rational agreement behind the veil cannot justify moral principles as impartial. Self-interested choice cannot be a justificatory device for moral principles. While Rawls's form of contractualism falls under this criticism, Dworkin's use of the veil of ignorance does not. For this we have to compare the role that the veil of ignorance plays in justifying theories for Rawls and Dworkin. According to Rawls, agreement in the original position is what justifies moral principles. But for Dworkin, the veil of ignorance does not play any justificatory role. The veil of ignorance appears in the presentation of equality of resources. It is a device

³⁰ Scanlon, *What We Owe to Each Other*, pp. 189-97, also pp. 17-33 where Scanlon gives a different, more minimal, account of rationality.

³¹ Scanlon, "Contractualism and utilitarianism," pp. 119-28.

that illustrates and specifies how equality of resources works. Dworkin identifies problems for a conception of distributive equality. He identifies the challenge that handicaps and differences in talents pose. He gives independent arguments why a theory of equality should be ambition-sensitive but avoid endowment-sensitivity. The veil of ignorance helps specifying what this means in particular circumstances. It is more a component of the theory itself than a justificatory device for the theory.

Equality of resources is justified in other ways. Part of its justification stems from the criticism of its competitor, equality of welfare.³² Part of its justification stems from the presentation alone, as an attractive conception of equality. Part of its justification stems from its ability to incorporate concerns of individual responsibility and liberty. In later works Dworkin further develops this holistic justification. Equality of resources is the conception of equality which meets two desiderata for a theory of distributive justice. It blends a model of how governments show equal concern for the fate of their citizens with a model of how government show equal respect for the responsibility of each citizen to live well. The requirements of equal concern and equal respect are requirements for governments. They embody principles of how to treat others, but they receive their force in turn from principles of how to live well and treat one's own life. In order to live well we need to acknowledge the importance of living well and we need to insist and make use of our responsibility to live our life well. Equality of resources is thus justified holistically as the conception of equality that not only integrates political values like equality, responsibility, and liberty, but also coheres well with a model of one's duties to oneself.³³

Nonetheless, Dworkin's equality of resources which addresses the question of distributing resources between persons makes the answer to this question dependent on the choices in the hypothetical insurance market and of judgments of intra-personal prudent choice behind the veil of ignorance. Does this not transform

³² Dworkin, *Sovereign Virtue*, ch. 1.

³³ Ronald Dworkin, *Justice for Hedgehogs* (Cambridge, MA.: The Belknap Press of Harvard University Press, 2011), pp. 1-15, 191-218 (for his account of personal dignity and living well), pp. 351-63 (for how equality of resources fits in this structure). Already in *Sovereign Virtue* Dworkin's theory of equality is connected to questions of liberty (ch. 3) and questions of the good life (ch. 6).

such inter-personal trade-offs to intra-personal trade-offs?³⁴ For any given transfer, we can sensibly say that the transfer is permissible or impermissible depending on the choice behind the veil of ignorance. Dworkin, for example, holds that it would be unfair to make expensive medical insurance compulsory that extensively covers end of life care for the last few months of one's life. Since only few people would heavily insure themselves against such risk, a mandatory transfer would be overreaching.³⁵ This points to an ambiguity in the justificatory version of the separateness of persons objection. Should we interpret it narrowly as a constraint on the reasons in favor of a moral theory? If so, Dworkin's argument would not violate the separateness of persons. His reasons for accepting the insurance test do not depend on any self-interested choice. Or should we interpret it widely as a constraint on moral reasons in favor of actions, including those which the moral theory provides? If so, Dworkin's argument would violate the separateness of persons. The insurance test makes intra-personal trade-offs pertinent to the question of inter-personal trade-offs.

The alleged violation of the separateness of persons is that it turns different people's lives into mere possibilities of a single person's life. The key criticism of this model of justification is voiced by Nagel. Nagel criticizes that it is very different whether a bad life is a mere possibility which may not materialize, or whether some person has to lead this bad life no matter what.³⁶ The problem is that it is inappropriate and impermissible to treat bad outcomes as mere possibilities when they will be actually realized. This confuses bad actualities, i.e. real suffering, with bad eventualities. Choosing a distribution for society is, therefore, different from choosing a risk profile for oneself. In Harsanyi's model this is very clear. Harsanyi tells us to treat all actual outcomes as if they were possible outcomes for one chooser.

Dworkin's justification is different, however. He provides an argument for what constitutes a fair share of resources in just circumstances. The veil of ignorance thought experiment determines that in just circumstances individuals would not have insurance for such medical care. So why, we may ask, should anyone complain

³⁴ Marc Fleurbaey, "Equality of Resources Revisited," *Ethics* 113 (2002): 82-105, at p. 90.

³⁵ Dworkin, *Sovereign Virtue*, pp. 314-15.

³⁶ Nagel, *The Possibility of Altruism*, pp. 138-39.

if they are made as well off as they would be in circumstances of justice?³⁷ The deep justification for using the insurance test does not rely on the idea that we can write off bad eventualities. Even the insurance test itself is disanalogous to Harsanyi's argument. In Harsanyi's case, the decision is about a distribution of goods across individuals. In Dworkin's case, the decision is about individual entitlements. The insurance test does not serve to select entire distributions.

The example Nagel uses to illustrate his objection further shows the difference between the two models. Nagel writes that it could be rational to take a small risk of enslavement in exchange for a good chance of opulent luxury.³⁸ Harsanyi's justification has to accept this trade-off. Dworkin's justification does not. In just circumstances everyone would be ensured never to be enslaved. No understanding of equal concern and respect for each person's life would accept enslavement. The prospect of enslavement is not just treated as a bad eventuality here. This indicates that the justificatory objection should be read more narrowly. The separateness of persons is a constraint on the reasons given for a moral theory, not a constraint on the questions a moral theory may ask.

This shows that Scanlon's criticism does not extend to all versions of rational choice behind the veil of ignorance. It applies only where rational choice behind the veil is part of the justification rather than where it is a tool in specifying a conception of an abstract moral ideal. This leads to an intriguing question about Rawls's use of the veil of ignorance. While Rawls presents agreement in the original position as providing an argument for his principles of justice, there is an alternative reading to Rawls's use of the original position. In an article on the original position Dworkin interprets the original position not as the foundations of Rawls's theory of justice but rather as a component of it.³⁹ For Dworkin the original position is a midway point in the justification of principles of justice. The original position, he argues, is a device

³⁷ The force of this question is increased by Dworkin's other arguments. For example, his criticism of equality of welfare, if successful, refutes the reply that only actual welfare matters. His arguments for incorporating personal responsibility, if successful, strengthen the case that we could not complain in just circumstances if our option luck on the insurance bet has been bad.

³⁸ Nagel, *The Possibility of Altruism*, pp. 138-39.

³⁹ Ronald Dworkin, *Taking Rights Seriously* (London: Duckworth, 1978), ch. 6.

that is justified by a deeper theory of political rights that accords every member of society with a right to equal concern and respect. The original position is then an appropriate device to test different conceptions of equal concern and respect. This is because the original position gives every party a veto power that corresponds to the political right of equal concern and respect. Rejection in the original position shows that the proposed basic structure of society violates the right to equal concern and respect. Principles of justice are ultimately correct not because they would attain hypothetical consent, but because they would be the best conception of a more fundamental right to equal concern and respect.⁴⁰ Interpreted in this way, does Rawls's original position still fall foul to Scanlon's criticism?

The alternative reading of the original position does not hold that hypothetical agreement is of moral importance by itself. Nor does it aim to give an expression of impartiality by the model of self-interested choice. Instead, the choice in the original position reflects the veto power everyone has in virtue of fundamental political rights. Scanlon's criticism seems to be less forceful for these reasons. The alternative reading does not reduce justifiability to self-interested choice. The conditions of the original position set the boundaries for the political right. Dworkin gives the example of Hobbes's state of nature. He assumes that Hobbes's deep theory is a right to life and that this explains why the parties in Hobbes's state of nature value security to an extreme extent. Of course, this reduction of political rights to self-interested choice may fail. But whether or not it fails, it is not subject to the criticism raised by Scanlon.

While this vindicates the device of the original position in general, it does not vindicate Rawls's adaptation of it. The choice in the original position is still made with assumptions that betray the separateness of persons. The different role that the original position plays in a larger moral theory does not take away the fact that there is no sense we can attach to the separateness of the parties of the original position.

⁴⁰ Rawls reply to Dworkin gives a different answer. Rawls's answer shares with Dworkin's reinterpretation the idea that the original position is a device to give content to a more basic moral notion. In Rawls's reply he highlights that the original position specifies what fair terms of cooperation between free and equal citizens look like. John Rawls, "Justice as Fairness: Political not Metaphysical," *Philosophy & Public Affairs* 14 (1985): 223-51, at pp. 234-39.

This feature of the original position also belies a key motivation Dworkin identifies for using the social contract device. Dworkin argues that the social contract, or hypothetical consent, can be motivated by a rights-based theory because it gives each distinct individual a veto power over political institutions.⁴¹ The veto power is an exercise of their fundamental political rights. It is therefore limited by the scope of these rights. The veil of ignorance imposes limits on what the parties in the original position can veto. It therefore limits their veto rights. This can be interpreted as a reflection of the scope of their political rights. The problem is that Rawls's veil limits knowledge in a way that not only limits the ability to veto. It also makes disagreement impossible. Any alternative to Rawls's two principles would be vetoed by every party of the original position. This removes the original motivation that each *distinct* individual has a veto power. A veil of ignorance, like Rawls's, that reduces the parties of the original position to a single or only few types thereby crosses a line that other social contract devices do not.

F. *From Rawls to Dworkin: Collective Assets*

One further issue remains. In two passages of *A Theory of Justice* Rawls indicates that the difference principle constitutes an agreement about how to distribute and share the benefits of natural talents as collective assets.⁴² But is this not saying that natural talents belong to all of us as a whole? Does this not go against the separateness of persons? If talents are part of one's person, then why are we allowed to treat them as collective assets? Talents are often central to one's sense of self, so any distinction of talents from person would seem forced and highly artificial.⁴³

Whether or not we regard talents as collective assets depends on our justification for the difference principle. Not every justification for redistribution needs to assume that talents are assets that are collectively owned. If this is so, then

⁴¹ Dworkin, *Taking Rights Seriously*, pp. 176-77.

⁴² Rawls, *A Theory of Justice*, rev. edn., pp. 87, 156.

⁴³ This complaint is first raised by Robert Nozick. Nozick, *Anarchy, State, and Utopia*, pp. 228-29. See also Michael Sandel, *Liberalism and the Limits of Justice*, 2nd edn. (Cambridge: Cambridge University Press, 1998), pp. 77-81.

the objection to collective assets is a justificatory version of the separateness of persons objection and not a substantive one.

There are four main ways in which principles of distributive justice could be justified. Distributive justice may be justified *derivatively*. In this case, the justification would not make any appeal to distribution at all. For example, we might justify distributive justice only to the extent that is necessary to ensure democratic stability. This kind of argument makes no appeal to the idea of collective assets. It only holds that the moral reasons for democratic stability are strong enough to require redistribution. The second justification is by treating principles of distributive justice as principles for the *division of surplus of mutually beneficial cooperation*. Rawls's appeal to collective assets is surprising given how important the idea of dividing the surplus of reciprocal cooperation is to his theory. The division of surplus assumes collective work, i.e. cooperation, instead of collective assets. The third justification is *compensatory*. Redistribution is done in order to compensate individuals for unfair disadvantage. This justification does not need to assume that talents are collective assets. Instead, the unfortunate are simply compensated. Only the last and fourth way assumes collective assets. I call this approach *aggregate and divide*. Given a fixed currency of justice, we determine the total of this currency. We then divide up the currency into individual shares according to the formula that our theory of justice provides. Utilitarianism can be explained in this manner. Utilitarians simply add up all welfare and divide it in order to bring about the distribution that maximizes the sum total of welfare. This justification has to assume that talents are collective assets.⁴⁴

The difference between the last two justifications can be seen in Dworkin's construction of equality of resources. Dworkin contrasts his hypothetical insurance market with an alternative approach to the problem of handicaps.⁴⁵ He suggests that a person's "physical and mental powers" might count as resources for the purpose of the initial auction division. He even concedes that these powers are indeed personal

⁴⁴ This also holds for Harsanyi's justification. Harsanyi simply assumes that principles of distributive justice have to be chosen. There is no attempt in justifying his principles as a form of compensation or as a division of cooperative surplus or by way of some derivative justification.

⁴⁵ Dworkin, *Sovereign Virtue*, pp. 79-80.

resources. In this alternative scenario every person would first receive a compensatory share of external resources before the auction can proceed with the remaining external resources.

Dworkin rejects this approach on the following grounds.⁴⁶ First, an initial compensation would require a standard of normal powers which is difficult to give. Second, it may not provide an upper bound for compensation. In practice compensation would then need to be determined by the political process. Third, he objects that treating resources this way would amount to seeing them as transferable and fungible between persons. The first two arguments are unconvincing. Dworkin concedes that mental and physical powers are resources. It should be possible then to determine how valuable these resources are. This might be very complex for any human to do. But this is in effect an epistemic concern. It does not establish that there is no right answer to the question of how valuable one's personal resources are and how much initial compensation one is owed. Similar remarks hold for the second argument. Dworkin points out problems of implementation. Politicians might be unwilling to transfer even more resources to those badly off, but it may still be the case that this is what justice requires.

The third argument is more promising. Dworkin writes that personal resources "cannot be manipulated or transferred, even so far as technology might permit".⁴⁷ Dworkin does not give a reason why transfers are impermissible even when they are feasible. Dworkin's remark makes sense, however, if it is interpreted as a demand to respect the separateness of persons. It is incompatible to respect the separateness of persons while treating personal resources as transferable and fungible. This would amount to the divide and aggregate approach to distributive justice which I outlined above. We can then see that personal resources are excluded in the initial distribution in order to respect the separateness of persons. This creates the need for Dworkin's alternative solution, the hypothetical insurance market, and thereby for the veil of ignorance. The veil of ignorance is only introduced in Dworkin's theory in order to respect the separateness of persons.

⁴⁶ Dworkin, *Sovereign Virtue*, pp. 79-80.

⁴⁷ Dworkin, *Sovereign Virtue*, p. 80.

Having distinguished the various modes of justification for principles of justice, we can see that Rawls as well can be defended against the problem of collective assets. Only *aggregate and divide* makes an appeal to collective assets which violates the separateness of persons. The other three approaches do not need to regard talents as collective assets.⁴⁸ Rawls's theory has many resources to develop arguments based on all three permissible strategies, even though I will not pursue this task here.

My interpretation can, however, explain the revisions Rawls made for the revised edition of *A Theory of Justice* on this point. In the original edition of *A Theory of Justice* Rawls writes that “[we] see then that the difference principle represents, in effect, an agreement to regard the distribution of natural talents as a common asset and to share in the benefits of this distribution whatever it turns out to be.”⁴⁹ In the revised edition this passage is removed. Rawls there only writes that “[the] two principles are *equivalent* ... to an undertaking to regard the distribution of natural abilities in some respects as a collective asset I do not say that the parties are moved by the ethical propriety of this idea.”⁵⁰ In the later formulation Rawls distances himself from reasoning that treats natural abilities as common assets and as such disposable by everyone. He expressly says that the parties are not moved by this ideal. This inclusion and the other modifications indicate that Rawls himself is not moved by this ideal either. But he does not distance himself from the two principles of justice. Rather, he points out that the principles of justice which can be justified independently are extensionally equivalent to principles justified by appeal to collective assets.

My interpretation can also shed light on the emphasis that the *distribution* of natural talents is regarded as a common asset as opposed to the individual talents

⁴⁸ This result is not trivial. Equality of outcome, for example, seems much less plausible if it has to be justified on either of the three grounds. It is difficult to see why every inequality should have to be compensated for. It is also difficult to see why the division of surplus should lead to flat equality irrespective of different contributions and different non-cooperation baselines. Lastly, it is difficult to see why non-distributive ideals would require such a demanding distributive implementation.

⁴⁹ Rawls, *A Theory of Justice*, p. 179.

⁵⁰ Rawls, *A Theory of Justice*, rev. edn., p. 156. Emphasis added.

themselves.⁵¹ Regarding the distribution of talents as an asset means that we regard the assembly of different talents as one asset of the community as a whole. Yet, the distribution of natural talents can also be interpreted in light of the aggregate and divide approach. Under this it is the totality of talents that is an asset to the society as a collective. But this is not the most charitable reading of Rawls and one that he himself disavows.⁵² Instead, Rawls thinks that the distribution of natural talents refers to their complementarity. The distribution of talents leads to a division of labor that is part of a system of mutual cooperation. The division of labor and mutual cooperation itself then is a collective asset. This can be captured well in my taxonomy as an approach that regard the difference principle as a principle for the division of mutually beneficial surplus.

G. *Summary*

My discussion of the justificatory requirement of the separateness of persons objection has shown that it is possible to employ the veil of ignorance without disrespecting the separateness of persons. For this, three things have to be kept in mind. First, the veil of ignorance must be designed in a manner that allows us to distinguish between the choosers behind the veil. Second, self-interested choice behind the veil cannot justify moral principles. The veil of ignorance must either be justified by some deeper principle or it must be used in presenting the theory as opposed to justifying it. Third, the veil of ignorance cannot be used as a tool for diving up collective assets.

⁵¹ See Samuel Freeman, *Justice and the Social Contract* (Oxford: Oxford University Press, 2007), pp. 115-19; and John Rawls, *Justice as Fairness. A Restatement* (Cambridge, MA.: Harvard University Press, 2001), pp. 74-77.

⁵² Rawls, *Justice as Fairness. A Restatement*, pp. 74-77.

III. Principles for Separate Persons

A. Harsanyi's Average Utilitarianism

For the remainder of the chapter I will test the suggestion that the veil of ignorance systematically selects those principles of justice which substantively violate the separateness of persons. In particular, the concern is that the principles will treat intra-personal and inter-personal trade-offs the same and do not respect what I shall call the Shift between these different kinds of trade-offs. I begin with Harsanyi.

Harsanyi's average utilitarianism violates the separateness of persons substantively. This is the case regardless of which interpretation of Harsanyi's view is taken. Harsanyi himself argued that his veil of ignorance will lead to the acceptance of average utilitarianism, but there is sustained criticism against this interpretation.⁵³ As it turns out, on either interpretation of Harsanyi's veil, his result violates the separateness of persons. Average utilitarianism does not respect the Shift. For inter-personal choices average utilitarianism endorses the choice which maximizes average well-being. For intra-personal choices average utilitarianism endorses the choice which maximizes expected well-being. The intra-personal choice is then only a risky correlate of the inter-personal choice. In effect, the same principle of choice is used in both circumstances.

The criticism of Harsanyi's interpretation is the following. Harsanyi's theorem relies on von Neumann-Morgenstern (vNM) utilities. Harsanyi interprets the vNM utilities to represent (or be a measure of) well-being. VNM utilities have two features which make this interpretation potentially problematic. First, they are not uniquely defined. There are an infinite number of mathematical transformations of one's utility function that are all equally acceptable as vNM representations. This

⁵³ Amartya Sen, "Welfare Inequalities and Rawlsian Axiomatics," *Theory and Decision* 7 (1976): 243-62; Amartya Sen, "Non-Linear Social Welfare Functions: A Reply to Professor Harsanyi," in *Foundational Problems in the Special Sciences*, ed. Robert E. Butts and Jaakko Hintikka (Dordrecht: D. Reidel, 1977), pp. 297-302; John A. Weymark, "A reconsideration of the Harsanyi-Sen debate on utilitarianism," in Elster, Roemer, *Interpersonal Comparisons of Well-Being*, pp. 255-320; and Roemer, "Harsanyi's Impartial Observer is not a Utilitarian".

means that inter-personal comparability of well-being might not be guaranteed. A further assumption of comparability is needed. All utility functions need to be scaled together to ensure comparability. Otherwise the function that represents “average utilitarianism” would be subject to arbitrary factors such as the specific mathematical representation used for each individual.

Second, vNM representations are risk-neutral with regard to utility. Imagine you are offered a gamble. A fair coin is flipped, and you receive either £100 or nothing. There will be some amount of money that you would rather receive for certain which makes you indifferent between the certain money or the gamble, say £45. Now you lead your life and experience a year with bad fortune. This makes it vivid for you how much you would like your life not to depend on pure luck. As a result, you become more risk-averse than you were before. But the year does not change your attitude towards money or the benefits you draw from money. After this year I ask you to play the game again. This time you would be indifferent between the gamble and a lesser amount of money, say £40. According to the vNM measure, this means that your utility of receiving £45 is now higher than it was before.

But how is this possible? You did not change your evaluation of the outcome B. You value money just as much as you did before. What changed was your attitude towards risk. Psychologically there are two different reasons for why one might have changed one’s mind. It is possible for an agent to have changed their mind about the value of money. In this case the agent would consider having more money less valuable. Alternatively, the agent may have changed their attitude towards risk. In this case the agent would prefer to avoid risk without needing to think that the possible gains are less valuable. The agent would simply dislike betting.⁵⁴ The von Neumann-Morgenstern framework treats these two psychological explanations the same. “Risk aversion” in the von Neumann-Morgenstern framework simply means having diminishing marginal utility with regard to some good. This collapses the two

⁵⁴ See also J.W.N. Watkins, “Towards a Unified Decision Theory: A Non-Bayesian Approach,” in Butts and Hintikka, *Foundational Problems in the Special Sciences*, pp. 345-79, at pp. 368-75; Lara Buchak, *Risk and Rationality* (Oxford: Oxford University Press, 2013), pp. 24-36; and H. Orri Stefánsson and Richard Bradley, “What is Risk Aversion?,” *British Journal for the Philosophy of Science* 70 (2019): 77-102, at pp. 80-83.

different kinds of psychological attitudes into one measure. After changing one's mind, one's utility of B has increased.

But if the two attitudes are really distinct, then your well-being has not changed in my example. Well-being refers to an agent evaluation of how well one's life goes, it does not make reference to a person's attitudes on risk or gambling. As long as it is rationally permissible for an agent to be risk-neutral with regard to their good, this means that the vNM measure does not measure well-being for all agents. If we used the vNM measure as a guide to the distribution of resources we would, in the words of Kenneth Arrow, make giving of benefits dependent on the tastes of individuals for gambling.⁵⁵

A better interpretation of vNM utilities is that they do not represent well-being but rather an index that includes both one's valuation of states of affairs (i.e. one's well-being) *and* one's attitudes towards risk. Average utilitarianism on the other hand would require us to maximize the average of well-being as opposed to the average of this index. Harsanyi's argument would then justify only maximizing average vNM utilities. Maximizing average vNM utilities, however, does not respect the Shift. VNM utilities are constructed by accounting for rational intra-personal trade-offs. As a moral theory, maximizing average vNM utilities would simply use the same kind of mechanism for inter-personal trade-offs. Whichever way we interpret Harsanyi's theorem, it cannot account for the Shift and therefore Harsanyi's veil does not provide a moral theory that respects the separateness of persons.

B. Rawls's Two Principles of Justice

Rawls avoids the substantive version of the separateness of persons objection by restricting his principles of justice. They apply only to the basic structure of society. This immunizes Rawls from any possible violation of the Shift. His principles are simply not meant to apply to individual, intra-personal trade-offs. Furthermore, the restriction is not ad hoc. His conception of justice as being the first virtue of social

⁵⁵ Kenneth J. Arrow, *Social Choice and Individual Values*, 2nd edn. (New York: John Wiley & Sons, 1963), pp. 9-10.

institutions is internally coherent and internally motivated. This point is important. It is technically possible for every moral principle that seemingly violates the Shift to limit its applicability to inter-personal choices. If every such limitation is permitted, then the Shift cannot be an effective argument against moral principles. In some cases, such limitations will be internally incoherent, they will belie the motivation given for the principle in the first place. This is clearly not the case with Rawls's two principles of justice.

C. *Dworkin's Equality of Resources*

Dworkin's equality of resources makes decisions about social transfers dependent on individual hypothetical insurance decisions. Since insurance decisions balance out risks between different possible futures there is a concern that the theory assimilates inter-personal trade-offs to intra-personal trade-offs. To assess this objection, it is worthwhile considering a concrete case. Alex Voorhoeve has devised one against the application of equality of resources to health care rationing.⁵⁶ Voorhoeve imagines a three person society which can choose between three health care insurance plans. Unhealthy would most benefit from a large insurance policy, while Healthy would most benefit from a small insurance policy. Avy, the third member, knows that she will either develop the condition that Unhealthy has or have the health status of Healthy. The best option for her is a medium insurance plan. The hypothetical insurance model asks us to determine which insurance Unhealthy and Healthy would have purchased had they been unaware of their condition. Avy is currently unaware and chooses a medium plan. If we accept the judgment of a representative individual, then we would follow Avy's judgment and select a medium plan for Unhealthy and Healthy too.

⁵⁶ Alex Voorhoeve, "May a Government Mandate more Comprehensive Health Insurance than Citizens Want for Themselves?," in *Oxford Studies in Political Philosophy. Volume 4*, ed. David Sobel, Peter Vallentyne, and Steven Wall (Oxford: Oxford University Press, 2018), pp. 167-91, at pp. 172-74. For the criticism in general see John E. Roemer, "Equality of Talent," *Economics & Philosophy* 1 (1985): 151-88; and Fleurbaey, "Equality of Resources Revisited," pp. 90-97.

This brings out the separateness of persons objection. The trade-off between Unhealthy and Healthy is resolved in effect by the intra-personal trade-off that Avy faces. The hypothetical insurance model does not respect, so it seems, the Shift. Inter- and intra-personal trade-offs are treated the same.

The example Voorhoeve uses is directed against the “representative prudent individual test” (RPIT) for health care coverage. RPIT is a simplification of equality of resources for health care. Once this simplification is removed, we can see that equality of resources can evade the separateness of persons objection. Under equality of resources there is not one choice of a representative individual which determines the overall distribution of resources. Instead, it is the interplay, in a market, of various person’s choices behind a veil of ignorance which determines the distribution of resources. It is not a single hypothetical insurance decision, but rather the result of a hypothetical insurance market that determines the right amount of transfers to those unfairly disadvantaged.

In Voorhoeve’s example RPIT does not respect the Shift because of restrictions in the thought experiment. The size of the society is small, and all members of the society share an attitude towards risk. His example also uses well-being as a currency of justice, which is incompatible with the model of equality of resources. The restrictions are crucial to bring out one example in which equality of resources coincides with a theory that decides inter-personal trade-offs the same way as intra-personal trade-offs. This is different from the charge that equality of resources turns questions of inter-personal transfers into questions of intra-personal risk-taking. Utilitarianism, for example, *systematically* violates the difference between inter-personal and intra-personal trade-offs. To see why one example in which the answers to an inter-personal and intra-personal trade-off coincide is insufficient, consider the following. Under special circumstances utilitarianism coincides with outcome welfare egalitarianism in its allocation of resources.⁵⁷ It would be incorrect, however, to take this as evidence for the egalitarian character of utilitarianism.⁵⁸ It would also

⁵⁷ Namely, all individuals have the same preferences, there is diminishing marginal utility and distribution is costless.

⁵⁸ Utilitarianism is possibly egalitarian in a *different* sense, namely by embodying the principle that everyone counts for one and no one counts for more than one.

be incorrect to hold that for this reason outcome welfare egalitarianism violates the separateness of persons. It is noteworthy that both in the case of the egalitarian outcomes of utilitarianism and in Voorhoeve's application of equality of resources, the restrictions on the thought experiment are overwhelmingly unlikely to arise in the scenarios that the theory is designed for. Equality of resources is designed to answer the question of distributive justice for societies or governments, not for smaller units like families. Furthermore, the mere fact that equality of resources has the implication Voorhoeve shows it has in this case is not good enough reason, on grounds of counterintuitive consequences, to reject equality of resources. The argument would only be successful if it could be shown that equality of resources violates the Shift, but this, I argued, is not the conclusion we should draw from his case.

Dworkin does, however, introduce simplifying assumptions into his model of equality of resources which seem to make it appropriate to simply ask what the average member would have chosen.⁵⁹ This additional simplification would mean that equality of resources systemically violates the Shift. Can we improve on Dworkin's suggestion? Is there a simplifying assumption that respects the Shift? In a different context Lara Buchak has suggested that there is. Buchak argues that there is a wide array of reasonable and rationally permissible risk attitudes. When acting on behalf of others we should take a more conservative risk attitude, namely the most risk averse attitude which is still reasonable. This default is the default that should guide our deliberation behind the veil of ignorance.⁶⁰ This model, Buchak contends, respects the separateness of persons. Inter-personal trade-offs should be decided in light of the default risk attitude. But individuals are free to depart from the default when it comes to intra-personal trade-offs. Indeed, since hardly anyone will adopt the most risk averse reasonable attitude, almost everyone will take different gambles.⁶¹

⁵⁹ Dworkin, *Sovereign Virtue*, pp. 78-79, 94-95.

⁶⁰ Buchak, "Taking Risks Behind the Veil of Ignorance," pp. 624-33.

⁶¹ Buchak, "Taking Risks Behind the Veil of Ignorance," pp. 640-42.

Buchak has given a way how inter-personal and intra-personal trade-offs can be differentiated even with a single simplifying assumption for inter-personal trade-offs. This assumption will not, however, help equality of resources. The project of equality of resources is to determine how people would be endowed in resources in just circumstances. The veil of ignorance is introduced to determine which resource bundle people would have chosen in these just circumstances. This question is one of individual choice and individual responsibility. Buchak's risk attitude only makes sense as an attitude of acting on behalf of others. Therefore, her suggestion cannot be integrated into equality of resources.

While Dworkin uses averaging as a simplifying assumption, he also makes clear that it is a second-best assumption.⁶² The model of equality of resources should always be refined with more information insofar as this information is available. Dworkin underestimates here the importance of this additional information, and he does so in a way that creates trouble for his own theory. Dworkin adjusts his veil of ignorance from Rawls in order to allow for the separateness of persons behind the veil. Resorting now to a standard of the average person risks repeating Rawls's mistake. The parties behind the veil of ignorance are still formally separate but what counts is only how the average member would vote. The result is that intra-personal and inter-personal trade-offs are treated alike, as they are in the counterexample to RPIT. Instead, equality of resources requires for its viability a larger informational basis that allows us to refine the details of the hypothetical insurance market.

This does not constitute an argument in favor of equality of resources. Equality of resources may still fail. There might be reasons apart from the separateness of persons that speak against it. There is also the possibility that equality of resources is hopeless if it cannot resort to simplifying assumptions of how the average person would choose behind the veil. If so, then the separateness of persons would play a crucial part in undermining the viability of Dworkin's theory of justice. Intriguing as this suggestion is, I cannot discuss it here.

⁶² Dworkin, *Sovereign Virtue*, p. 78 and 78fn5

IV. Conclusion

This concludes my discussion of the veil of ignorance. The general argument against the veil of ignorance is central in assessing its usefulness as a philosophical tool. But I have argued that the veil of ignorance can be used in a manner that respects the separateness of persons. Ronald Dworkin's use of the veil is one such example.

Three conditions need to be met for the veil of ignorance to avoid the justificatory version of the separateness of persons objection. First, the choice behind the veil must be a choice of distinct individuals. Depriving the parties behind the veil of information that allows us to differentiate between them turns the choice behind the veil into a choice of a single or only few representative individuals. Second, rational choice behind the veil cannot be the justification for a moral principle. The veil can play an important part as a component of the theory specifying its contents. Alternatively, the veil can be motivated by some deeper justification which makes self-interest behind the veil morally relevant. Third, redistribution cannot be justified on grounds that innate talents are collective assets. The veil of ignorance cannot be employed to divide a common pool of innate talents. Instead, it needs to be justified as part of a theory of compensation for unfair disadvantage, as part of a theory of division of the surplus of mutually beneficial cooperation, or derivatively by appeal to some other moral ideal.

Perhaps surprisingly, the veil of ignorance can violate the separateness of persons substantively. One way to safeguard against this is Rawls's. Rawls insulates the veil of ignorance by limiting its role to choosing principles for the basic structure of society. This way the resulting principles of justice cannot violate the Shift since they have no applicability to intra-personal choices. Another way is Dworkin's. Dworkin's equality of resources determines inter-personal trade-offs by the interplay of various different intra-personal decisions. As a result, the trade-offs will differ and respect the difference between the separateness of persons and the unity of the individual.

The separateness of persons is therefore not opposed to the veil of ignorance. Indeed, the veil of ignorance can sometimes be a tool to avoid violating the

separateness of persons. The limits that the separateness of persons sets to the veil of ignorance tell us however to depart from the simple model of individual rational choice of principles of justice behind the veil.

Chapter 3. Contractualism, Complaints, and Risk

I. Contractualism and Risk

The previous chapter focused on the demand that the separateness of persons imposes to respect the difference between intra-personal and inter-personal trade-offs. Moral theories should respect the difference between the separateness of persons and the unity of the individual. In the following three chapters, I want to focus on the separateness of persons in a narrow sense. Utilitarianism aggregates all benefits and burdens of an action in order to decide whether or not the action is permissible. It thereby conflates the different standpoints of different individuals and treats all benefits and burdens an action produces as if they were the benefits and burdens of one entity or one system of ends.

This objection to the aggregative feature of utilitarianism has motivated non-consequentialists to propose conceptions of morality that are based on the competing claims or complaints that individuals can raise. Placing the commitment to individual claims or complaints at the heart of morality seems a promising route to ensure respect for the separateness of persons and the separateness of the standpoints of distinct individuals. The most systematic of these proposals is contractualism as developed by T.M. Scanlon. Scanlon argues that an act's rightness or wrongness depends on its justifiability to each. As a test for justifiability, Scanlon proposes that the permissibility of an act depends on whether it follows from a principle that no one can reasonably reject. An act is permissible only when no one can reasonably reject a principle that entails the permissibility of that act. One natural idea is that the individual with the largest complaint has most reason to reject a principle. It then appears that a principle can be reasonably rejected only when the largest complaint is larger than the complaint anyone else could bring forward against any alternative principle.¹

¹ See Scanlon, "Contractualism and utilitarianism"; and *What We Owe to Each Other*, ch. 5.

The individualistic foundations of contractualism have given rise to an opposite concern, namely that contractualist morality is unduly concerned with the fate of single individuals.² Recently, Scanlonian contractualism has received scrutiny for the way it deals with cases where risks, rather than certainties of harm and benefit, are at stake.³ My discussion in this chapter will focus on Scanlonian contractualism, but my conclusions may apply more widely to any moral theory that places the idea of justifiability and individual complaints or competing claims at the heart of morality.

The debate around contractualism and risk is typically framed as a debate between two opposing views. *Ex ante contractualism* is concerned with prospects while *ex post contractualism* is concerned with outcomes.⁴ I believe that this framing is unhelpful. What can it mean to say that a theory of risk impositions is concerned with outcomes when it is designed to provide guidance in cases where we are uncertain about the outcome? With the help of a sequence of thought experiments from Michael

² See Brink, "The Separateness of Persons, Distributive Norms, and Moral Theory" and also my discussion of anti-aggregation in the introduction to this dissertation.

³ See Sophia Reibetanz, "Contractualism and Aggregation," *Ethics* 108 (1998): 296-311; Elizabeth Ashford, "The Demandingness of Scanlon's Contractualism," *Ethics* 113 (2003): 273-302; James Lenman, "Contractualism and risk imposition," *Politics, Philosophy & Economics* 7 (2008): 99-122; Barbara H. Fried, "Can Contractualism Save Us from Aggregation?," *The Journal of Ethics* 16 (2012): 39-66; Aaron James, "Contractualism's (Not So) Slippery Slope," *Legal Theory* 18 (2012): 263-92; Marc Fleurbaey and Alex Voorhoeve, "Decide As You Would with Full Information!," in *Inequalities in Health*, ed. Nir Eyal, Samia A. Hurst, Ole F. Norheim, and Dan Wikler (Oxford: Oxford University Press, 2013), pp. 113-28; Johann Frick, "Uncertainty and Justifiability to Each Person. Response to Fleurbaey and Voorhoeve," in Eyal, Hurst, Norheim, and Wikler, *Inequalities in Health*, pp. 129-46; T.M. Scanlon, "Reply to Zofia Stemplowska," *Journal of Moral Philosophy* 10 (2013): 508-14; S.D. John, "Risk, Contractualism, and Rose's 'Prevention Paradox'," *Social Theory and Practice* 40 (2014): 28-50; Johann Frick, "Contractualism and Social Risk," *Philosophy & Public Affairs* 43 (2015): 175-223; Rahul Kumar, "Risking and Wronging," *Philosophy & Public Affairs* 43 (2015): 27-51; Michael Otsuka, "Risking Life and Limb: How to Discount Harms by Their Improbability," in *Identified versus Statistical Lives*, ed. I. Glenn Cohen, Norman Daniels, and Nir Eyal (Oxford: Oxford University Press, 2015), pp. 77-93; Joe Horton, "Aggregation, Complaints, and Risk," *Philosophy & Public Affairs* 45 (2017): 54-81; and Korbinian R ger, "On Ex Ante Contractualism," *Journal of Ethics and Social Philosophy* 13 (2018): 240-258.

⁴ For the former see James, "Contractualism's (Not So) Slippery Slope"; John, "Risk, Contractualism, and Rose's 'Prevention Paradox'"; Kumar, "Risking and Wronging"; and Frick, "Contractualism and Social Risk". For the latter see Fleurbaey and Voorhoeve, "Decide As You Would with Full Information!"; Otsuka, "Risking Life and Limb"; and R ger, "On Ex Ante Contractualism".

Otsuka, I provide a more helpful way of understanding what is at stake between different contractualist approaches to risk (Section II).⁵ In addition, the sequence allows me to propose a new view on contractualism and risk, which I call *objective ex ante contractualism* because of the special importance that it gives to objective as opposed to epistemic probability. My version of contractualism focuses on the complaints of would-be victims whose fate is already determined. After discussing the sequence, I will show that a natural extension of the sequence highlights that two principles which ex post contractualism should ideally fulfill are inconsistent with one another (Section III). In Section IV, I will present the defense of my objective ex ante view by arguing that it provides us with the best model of the key contractualist idea of acting in ways that are justifiable to each. Section V responds to objections.

II. Otsuka's Sequence

Dust. A comet is en route to the Midwestern United States carrying a pathogen that will soon lead to millions of people being infected and dying. The government is briefed on two alternative ways of containing the pathogen. The first option has the side effect that a different hazard will be released over Florida. It is known that it would cause Bob Johnson, a resident of Boca Raton, to lose one leg. Unfortunately, Bob Johnson cannot be evacuated in time. The second alternative has the side effect that the hazard will have to be released in a dust cloud over California. Each of 40 million Californians faces a small risk of death and it is known that exactly one Californian will die. The Californian who will die has a genetic predisposition which will cause his or her death upon being subjected to the dust.

Intuitively, the right course of action here would be to release the hazard over Florida and cause Bob Johnson to lose a leg. But it appears that contractualism struggles to explain this intuitive answer. Bob Johnson's complaint against choosing to release the hazard is not discounted. It is certain that he will suffer. The complaints

⁵ Otsuka, "Risking Life and Limb," pp. 77-88.

of the Californians should be discounted however. The likelihood of each of the 40 million Californians to be the one who dies is only 1 in 40 million. Although death is terrible, a 1 in 40 million chance of death is not altogether that terrible. We often incur similar risks when crossing the road, cooking with gas or swimming in the ocean. The complaint against the imposition of the risk of death would suddenly be a rather trivial moral complaint. How can such a trivial moral complaint outweigh the quite serious complaint of Bob of losing his leg?

One way for contractualism to accommodate the case is by pointing out that all the complaints combined add up to something significant: a complaint of the magnitude of certain death. But this response leads to highly counterintuitive results in other cases.

Jones. Jones, a worker in a TV transmitter room, has had an accident. He is now lying on the floor and suffering extremely painful electric shocks. There is only one way to save Jones, namely by interrupting the current transmission signal for about fifteen minutes. This in turn will cause millions of viewers to be upset who want to see the football World Cup match that is in progress.⁶

If we add up the complaints due to inconvenience and upset of all the millions of viewers, it seems that they will outweigh Jones's complaint against being subject to pain. But here it is clear that we should not let Jones suffer for the relatively mild loss of missing fifteen minutes of a football match. We should not aggregate morally trivial complaints so that they outweigh serious moral complaints of single individuals.

Otsuka, in his discussion of *Dust*, resists this solution and instead points to a different feature of the case. Unlike in Jones's case, in *Dust* there is one person who will experience grave harm. The aggregated complaints add up to the real-life predicament of one person. We do not need to imagine a social entity that experiences the harms of dying, but there is an individual made out of flesh and blood who will die. It is merely a fact concerning our informational limitations that prevents us from identifying that person in the same manner that we were able to identify Bob Johnson.

⁶ Scanlon, *What We Owe to Each Other*, p. 235.

Yet we can still say something about the individual who is going to die. The person who is going to die is “the Californian with the genetic predisposition”. The complaint of “the Californian with the genetic predisposition” is non-discounted. Her (or his) complaint would outweigh Bob Johnson’s complaint.

Now is the complaint of “the Californian with the genetic predisposition” a complaint *ex ante* or *ex post*? *Ex post* contractualism can account for this complaint. We know that the result of the action will be one person dying. Since the outcome distribution of the action is already known to us, an *ex post* contractualist can peek ahead, anticipate this distribution, and assign complaints to those affected by it.

But can *ex ante* contractualism? I think it can. “The Californian with the genetic predisposition” is a person with a determinate identity when we make the decision. Regardless of what happens and regardless of our action, “the Californian with the genetic predisposition” will always be the same person. If we limit our attention to only those possible worlds that are possible outcomes of our action, then we can say that “the Californian with the genetic predisposition” rigidly designates over this restricted domain of discourse. Since only those possible worlds that constitute possible outcomes of our actions are of interest to us, I will simply refer to such descriptions as “rigid designators”.⁷ Releasing the hazard over California will impose the certainty of death on this existing person with a determinate identity. From the *ex ante* perspective, “the Californian with the genetic predisposition” can object to the imposition of a 100 percent risk of death. We do not need to appeal to the outcome of the action *ex post* to make this claim.

This means that our understanding of *ex ante* contractualism should be broader. The classical version of *ex ante* contractualism focuses on the risks as faced by individuals with proper names or otherwise identifiable individuals. But not all versions of *ex ante* contractualism focus on these risks. The version of *ex ante* contractualism that I defend focuses on the complaints that rigidly designated

⁷ This definition also includes an element of temporality in the *ex ante*/*ex post* distinction. The possible worlds that are possible outcomes of the action are those possible worlds which coincide in their history until the point of action. Rigid designators are descriptions that refer to information that is contained in the shared history. Non-rigid designators are descriptions that refer to information about the future where the possible worlds no longer coincide.

individuals can raise. The two forms of ex ante contractualism differ thereby in whose complaints they focus on. This in turn is linked to a distinction between two kinds of risk: epistemic risks (credences) and objective risks (chances).⁸ The distinction that I am relying on here classifies some probability functions as expressing our uncertain degrees of belief or confidence about the world. These are epistemic probability functions, also called credence functions. By contrast objective probability functions express a mind-independent idea of probability. The objective probability function, a chance function, reflects information about the world and not about our knowledge of the world. If there are non-trivial objective probabilities, then there are truly “chancy” events. While there are various theories on what chances are, the differences between them are not important for my arguments.⁹ What I rely on is solely the contrast between chances and credences.

In *Dust* we only have epistemic probabilities for the risks that each identifiable Californian faces. However, we can give objective probabilities for the risk that “the Californian with the genetic predisposition” faces. This suggests an important link between the question of whose complaints we are interested in and what kind of risk we are interested in. By focusing on rigidly designated individuals, objective ex ante contractualism gives primacy to objective risk assessments over epistemic risk

⁸ I follow here the orthodox tradition in the philosophy of probability dating back to Rudolf Carnap who distinguished between two concepts of probability (frequentist and evidential) which are examples of the broader approaches of chance and credence. See Rudolf Carnap, “The Two Concepts of Probability,” *Philosophy and Phenomenological Research* 5 (1945): 513-32; Anthony Eagle, “Chance versus Randomness,” in *The Stanford Encyclopedia of Philosophy*. Spring 2019 Edition, ed. Edward N. Zalta (URL: <<https://plato.stanford.edu/archives/spr2019/entries/chance-randomness/>>), sec. 1; and Alan Hájek, “Interpretations of Probability,” in *The Stanford Encyclopedia of Philosophy*. Fall 2019 Edition, ed. Edward N. Zalta (URL: <<https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>>), sec. 3.

⁹ The most common approaches are frequentism, propensity views, and Best Systems Approaches. In addition, some philosophers embrace a “no theory” approach to chances according to which objective probabilities are not reducible to anything else like frequencies or propensities. For an overview see Hájek, “Interpretations of Probability”, for the no theory approach see Elliott Sober, “Evolutionary Theory and the Reality of Macro-Probabilities,” in *The Place of Probability in Science*, ed. Ellery Eells and J.H. Fetzer (Dordrecht: Springer, 2010), pp. 133-61, at pp. 148-54. An exception to my claim that my view on objective chance is independent between these views are actual frequentist views according to which objective probabilities only refer to actually occurring frequencies. Under such a view objective probabilities only represent statistical facts about reference groups and have no obvious moral significance.

assessments. Objective ex ante contractualism holds that in a case like *Dust* where the uncertainty is merely a matter of failing to identify the victim, we should choose descriptions that reveal the objective risks that individuals are facing. This is the “objective” component in objective ex ante contractualism.¹⁰

Let me move on to the next case in the sequence:

Wheel. The case is structurally similar to *Dust*. Again, we have a comet en route and a disaster about to occur. Again, one of our options is to release the hazard over Florida and cause Bob Johnson’s loss of a leg. But now our second option changes. As a side effect of averting the disaster, each Californian will be placed under a gigantic roulette wheel in the sky. The wheel will spin indeterministically and release a roulette ball that will kill exactly one person.

Otsuka reports his intuitive judgment that in *Wheel*, as in *Dust*, we should still prefer to release the hazard over Florida, causing the loss of Bob Johnson’s leg. But here we cannot rely anymore on the description of “the Californian who is genetically predisposed”. Instead, we would need to rely on a description like “the Californian who would be hit by the roulette ball” or “the Californian who would be most harmed by the decision”. These descriptions are non-rigid designators since different persons may die due to the falling ball. While the complaints of rigidly designated individuals have to be discounted, the complaints of non-rigidly designated individuals do not. The probability of someone being harmed by the wheel is 1. We can peek ahead and assign a complaint to that person. We may think that such statistical persons are still actual persons worthy of respect and with claims that ought to be taken into consideration.¹¹

¹⁰ Importantly the two kinds of risks are linked in a manner that should guard us from identifying epistemic or objective ex ante exclusively with one kind of risk. Whenever we have an objective probability for a given event (such as Charlotte Williams is going to be harmed), we should adjust our credence (i.e. our epistemic probability) to match the objective probability. The next case in the sequence is an example of this. This is, for example, an uncontroversial entailment of David Lewis’s Principal Principle. See David Lewis, “A Subjectivist’s Guide to Objective Chance,” in *Studies in Inductive Logic and Probability*, vol. 2, ed. Richard C. Jeffrey (Berkeley: University of California Press, 1980), pp. 263-93.

¹¹ See Norman Daniels, “Can There be Moral Force to Favoring an Identified over a Statistical Life?,” in Cohen, Daniels, and Eyal, *Identified versus Statistical Lives*, pp. 110-23, at p. 116; and Otsuka, “Risking Life and Limb,” pp. 85-86.

This cannot be reconciled with the ex ante perspective. The complaint of “the Californian most harmed by the decision” is not a complaint of any person with a determinate identity prior to the action. There is no token individual for whom it is true that she has imposed on her a 100 percent risk of death. Accordingly, my objective ex ante view holds that releasing the hazard over California is permissible in *Wheel*. Ex ante contractualism bases its complaints on the imposition of risk itself rather than on the eventual injurious outcome. This indicates that the important difference between ex ante and ex post concerns what the complaint is directed against, the risk itself or the eventual harm. A description like “the Californian most harmed by the decision” raises a complaint against the eventual harm. It reasons backwards from the eventual outcomes of the decision and bases complaints on these outcomes. This indicates a version of ex post contractualism.

Anticipating the strongest complaint ex post is easy in a case like *Wheel*. We know for certain how the benefits and burdens will be distributed in the outcome. We only lack information about who will be in which position. I now move on to a case where certainty about the resulting distribution is absent.

Guns. In this case we have the option to shoot down the comet with an automated weapons system. Unfortunately, the system also has guns in the sky pointed at each Californian. Each gun is operated by an indeterministic randomizer. The chance for each gun to fire and kill the person is 1 in 40 million. The guns, and thus the risks each gun imposes, operate independently of one another.

The objective risk for each Californian is the same as in *Wheel*, 1 in 40 million. Any assessment of rigid designators that relies on objective risks will be the same between *Wheel* and *Guns*. However, the assessment for non-rigid designators like “the Californian who will be most harmed” changes. Here we move away from certainty about the distribution that will come about and introduce risk as well. There is a 63 percent chance that at least one Californian will die, a 26 percent chance that at least two Californians will die, an 8 percent chance that at least three will die, and so on. What should ex post contractualists say about a case like this?

One answer is that *Guns* highlights the limits of ex post contractualism. Under this version of ex post contractualism we should draw a distinction between two types of cases. In some cases, like *Dust* or *Wheel* we know that the risk imposition will lead to harm while in *Guns* this is not guaranteed. Anticipating the complaint of the eventual victim is permitted in *Dust* and *Wheel* but not permitted in *Guns* according to this view. Since we do not know for certain that someone will be harmed, we cannot anticipate this complaint already.¹²

The problem with this version of ex post contractualism is that it relies on a distinction between risky cases that is morally dubious.¹³ Cases with guaranteed harms can easily be transformed into cases without guaranteed harm without changing anything of moral relevance. Take the example of a coin flip with inversely correlated harms and benefits. If the coin lands heads, A benefits and B is harmed. If the coin lands tails, A is harmed and B benefited. This is a case of guaranteed harm. Ex post contractualism would sometimes rule out this kind of risk even if it is in the antecedent interests of both. But what if the coin lands on the edge? This would be a freak accident, but nonetheless it is a possibility. Let us assume that no one will be harmed if the coin lands on the edge. The case is now one without guaranteed harm. If we are not allowed to anticipate any complaint ex post, we should do what is in the antecedent interests of both. Similar things hold for a version of *Wheel*. If we allow only a tiny chance that no one will be harmed, the restricted ex post view would allow the risk imposition since this case would no longer involve guaranteed harm. Yet if we are convinced that imposing the risk in *Wheel* is impermissible it should be impermissible even in this varied scenario. We need a different version of ex post contractualism.

Earlier I mentioned that in *Guns* we only know facts about what distributions of harms are to occur with which likelihood. For example, we know that the chance that at least one Californian will die is about 63 percent. One possibility for ex post

¹² Sophia Reibetanz Moreau defends such a view (“Contractualism and Aggregation,” pp. 302-4). Victor Tadros, in a different context, argues that these two kinds of risks are distinct (Victor Tadros, “Controlling Risk,” in *Prevention and the Limits of the Criminal Law*, ed. Andrew Ashworth, Lucia Zedner, and Patrick Tomlin (Oxford: Oxford University Press, 2013), pp. 133-55, at pp. 148-54).

¹³ Otsuka makes a similar point in “Risking Life and Limb,” p. 88.

contractualists is to translate these facts about distributions into complaints. Imagine we specify a ranking of all persons affected. The main ranking criterion is how strong each individual complaint against the action is. In cases where individuals are equally affected, we need other tie-breaking criteria. This way we can assign each individual a unique place in the ranking. Then we repeat this for all possible outcomes. We can now construct fictional characters or “statistical persons” based on these rankings. “The worst-off Californian” refers to the first-ranked person in each of the outcomes. “The second worst-off Californian” refers to the second-ranked person and so on. In cases of objective risk imposition, these designators are non-rigid since they refer to different individuals in different possible worlds. This construction allows us to assign unique complaints to individuals instead of being limited to talking about distributions of harms. Speaking of the complaints of non-rigidly designated persons brings the ex post perspective closer to the theoretical core of contractualism. It can provide a model of justifiability to each that an analysis of different distributions of harms cannot offer.¹⁴ Ex post contractualists should therefore accept the following principle.

Ex Post Discounting. When assessing the complaints of individuals, we should discount the complaints of non-rigidly designated individuals such as the worst-off, the second worst-off, and so on, by the improbability of harm.

As mentioned, in our case of *Guns*, this means that the complaint of the worst-off Californian is a discounted complaint against death rather than a non-discounted

¹⁴ Joe Horton has proposed a method that generates the same kinds of complaints. For Horton we should take the strongest complaint for each outcome and discount it by the probability that this outcome will come about. In the end we should aggregate all these complaints. Horton calls this the *expected strongest ex post complaint* (Horton, “Aggregation, Complaints, and Risk,” pp. 65-66). There is a subtle difference between my motivation and Horton’s motivation for why the ex post contractualist cares about these complaints. Horton’s motivation is that we should avoid states of the world in which strong complaints exist. My motivation is that we should give importance to the complaint of a person, albeit one that is non-rigidly designated. (This is similar to how Rawls gives importance to the claims of “the worst-off group” when justifying his difference principle.) I think that Horton’s motivation is a greater departure from the theoretical core of contractualism than mine. Contractualism distinguishes itself from other theories by focusing on the moral complaints of persons as opposed to placing emphasis on properties of states of the world as Horton’s motivation does.

complaint as in *Wheel*. The complaint is discounted by the 37 percent probability that the worst-off will not be harmed. But now the second worst-off Californian has a discounted complaint as well, as has the third worst-off, and so on. Should this difference matter?

Victor Tadros believes that it should. He gives the following argument based on an example that is a simpler version of the contrast between *Wheel* and *Guns*.¹⁵ Imagine we have two options. If we choose the first option, then it is guaranteed that one and exactly one person will die. If we choose the second option, then there is only a 75 percent chance that someone will die but there is also a 25 percent chance that two persons will die. Whatever we do, the risks to each rigidly designated individual are the same. Under one view the options are equally choiceworthy. If we choose the second option, there is a possibility that no one may die but this is balanced by the possibility that more than one may die. Tadros, however, argues that we should choose the second option because we have special reason to prevent a situation where harm will definitely occur. We should not regard the loss of two lives as twice as hard to justify than the loss of one life. This is because the two lives are separate and not part of one aggregate which suffers a double loss.

But it is hard to see why the separateness of persons should give us a special reason to avert definite harm. Tadros's argument implies that we have less reason to prevent an additional second death. Attaching special significance to the fact that harm will occur means attaching special significance to an isolate harm as opposed to a harm that occurs alongside many other harms. Yet deaths should have the same disvalue regardless of whether they are part of an action in which only one, two, or many people die. The death is just as tragic and severe for this person regardless of how many other people have died.¹⁶ Respect for each individual and for her separateness would seem to indicate that we should treat her loss by itself and not accord it more or less moral force because of the number of other people who have died. If this is true, then we should treat both options in Tadros's example as equally choiceworthy. The ex post contractualist should then regard *Guns* and *Wheel* as

¹⁵ Tadros, "Controlling Risk," pp. 153-54.

¹⁶ See also Otsuka, "Risking Life and Limb," pp. 88-92.

equally hard to justify. What should matter to us is the expected number of lives lost and not how the risk is distributed across non-rigid designators. This gives us a second principle that ex post contractualism would want to fulfill.

Equal Treatment for Equal Statistical Loss. We should treat cases alike if in both cases there is the same expectation of statistical loss and the only difference is the distribution of possible losses across possible outcomes.

III. A Problem for Ex Post Contractualism

Consider:

Gas. We receive yet another option to prevent the catastrophe. This time we have to release a gas in the air that will travel to California. Scientists tell us that there is the possibility that in California it will react by means of an indeterministic process with another substance and become toxic. If that happens, all Californians will die. However, they assure us that this is very unlikely. The objective probability of this occurring is only 1 in 40 million.

In one way, *Gas* is a continuation of *Wheel* and *Guns*. In all three cases, each rigidly designated Californian faces an objective risk of 1 in 40 million. The cases differ, however, in the distribution of risk across non-rigid designators. In *Wheel*, the distribution represents one extreme. All risk is concentrated in the likelihood of one person dying. In *Guns*, the distribution is spread out across all 40 million non-rigid designators ranked from “the worst-off” to “the best-off”. The risks for those higher up the list are very high, for those lower down the list they are minute. Now in *Gas* we face the opposite extreme. The risks are spread out perfectly even across all non-rigid designators. All non-rigid designators are tied, because whatever will happen, everyone in California shares the same fate. What is particularly interesting about *Gas* is that the distribution of discounted complaints is the same for rigid and non-rigid designators. Whether we use rigid or non-rigid designators to determine the justifiability of our action does not matter since both will yield the same result.

This is challenging for the ex post contractualist for the following reason: I have argued that ex post contractualists should accept the following two principles. They should accept *Ex Post Discounting*. This allows ex post contractualism to be applied to cases where harms are not guaranteed, and it provides the ex post perspective with a model of justifiability to each. Second, they should accept *Equal Treatment for Equal Statistical Loss*. This means that in *Wheel* and *Guns* what matters is the number of expected lives lost. The principle follows from accepting the claim that the disvalue of a given harm should not vary depending on how many other people will be harmed. The possibility that no person may die should be balanced by the possibility that more than one person may die.

My case *Gas* shows how these two principles can conflict. The number of expected lives lost in *Gas* is 1, just like in the other two cases. If *Wheel* and *Guns* are on a par, then so is *Gas*. But *Gas* contains only heavily discounted complaints by non-rigidly designated persons. This is because the complaint of the worst-off Californian is based on only a 1 in 40 million chance of death, a morally trivial complaint. Following *Ex Post Discounting*, it should be these discounted complaints that determine the justifiability of the risk imposition. If we want to follow *Equal Treatment for Equal Statistical Loss* and hold that the risk imposition in *Gas* is impermissible, we would need to aggregate the complaints in *Gas*. But whichever way we calculate the complaints, the complaints in *Gas* seem very close to the complaints by the many in *Jones*. The complaint of Bob Johnson resembles the complaint of Jones, the worker in the transmitter room. As it turns out, the strongest version of an ex post view leads to a case that is very much like *Jones*. If we allow aggregating the complaints in *Gas*, then why can't we aggregate the complaints in *Jones*?

One proposal is that while individual and non-aggregated complaints matter, aggregative considerations can determine whether it is reasonable to reject principles.¹⁷ Following this proposal, it is still individual complaints that matter. But

¹⁷ This is suggested by T.M. Scanlon as a general approach to aggregation in his latest revision of his contractualist views. Scanlon does not discuss risky cases in this context (T.M. Scanlon, "Contractualism and Justification," (unpublished manuscript)). Véronique Munoz-Dardé had earlier presented the idea that in some cases agents with strong complaints cannot reasonably reject principles. Munoz-Dardé invokes the idea of a threshold of reasonable demands that one can make on others. This allows for the possibility that a person with a stronger individual

their strength would be magnified by the number of people having the same complaint.

Ex Post Discounting (Multiplied). When assessing the complaints of individuals, we should discount the complaints of non-rigidly designated individuals such as the worst-off, the second worst-off, and so on, by the improbability of harm. The strength of their complaint is determined by multiplying the strength of their individual complaint with the number of non-rigidly designated individuals who will be equally affected.

According to this proposal it would be unreasonable for Bob Johnson to insist on his complaint given that there are so many complaints on the other side. The strength of the individual complaint opposing Bob Johnson is magnified by the number of people who would be similarly affected. Yet Jones is equally faced with many complaints on the other side. Why should we not be allowed to multiply the individual complaint of a single football fan by the number of football fans that are equally affected? If we are allowed to magnify this individual complaint, then it would be unreasonable for Jones to reject a principle which allows the World Cup match to be broadcasted. The proposal to allow individual and non-aggregated complaints to be amplified reintroduces aggregative reasoning through the backdoor. So what could distinguish between *Gas* and *Jones*? Why should we understand Bob Johnson's insistence on his individual complaint as unreasonable while Jones's insistence is reasonable?

Perhaps it is the following: In *Jones*, the small complaints stem from mere annoyance. In *Gas*, the small complaints are derivative of a very serious moral claim, namely the claim not to die. This very serious claim becomes less important to each individual taken separately due to the sharp discounting by the likelihood of its occurring. Maybe Bob Johnson's insistence is unreasonable while Jones's is not

complaint may not be able to reasonably reject a principle (Véronique Munoz-Dardé, "The Distribution of Numbers and the Comprehensiveness of Reasons," *Proceedings of the Aristotelian Society*, 105 (2005): 191-217, at pp. 208-15). I return to this proposal in Chapter 5 (*Aggregation, Balancing, and Respect for the Claims of Individuals*) and incorporate it in my positive proposal for the problem of aggregation.

because in Jones's case the opposing complaints are not complaints of the right kind. The trivial joy of watching football is not relevant to Jones's torture, while the risk of death, even if small, is relevant to Bob Johnson's lost leg. This proposal is coherent with what I wrote earlier about the opposition to aggregation. I wrote that "we should not *aggregate morally trivial complaints* so that they *outweigh serious moral complaints* of single individuals". Trivial complaints should not outweigh serious complaints regardless of the numbers involved. But this leaves open that complaints of similar magnitude or qualitative significance could outweigh each other depending on the numbers.¹⁸

In line with the earlier distinction between the complaints of the Californians and the complaints of the World Cup viewers, we could think of complaints as being qualitatively different for different levels of actual or possible harm. Following this idea, heavily discounting a complaint against being killed does not make this complaint morally trivial. The complaint is still qualitatively on a different level than the complaint against mere annoyance. This allows us to distinguish the aggregation in *Gas* from the aggregation in *Jones*.

One problem with the idea that risks of death are qualitatively different from very small certain harms is that the same answer is available to the ex ante contractualist. If we stop believing that heavily discounted risks of death are morally trivial, then we could engage in a limited form of aggregation in cases like *Wheel* too. And then ex ante contractualism can account for the same answer. In other words, once we adopt the view that heavily discounted harms are not morally trivial, we lose a key motivation for adopting ex post contractualism.

Second, treating risks of death as qualitatively different from small certain harms fails *Equal Treatment for Equal Statistical Loss* in a central case. It cannot treat

¹⁸ The idea that complaints can only be aggregated in some circumstances is called limited aggregation. The view is suggested by Scanlon, *What We Owe to Each Other*, pp. 238-41; and also endorsed and defended by Frances M. Kamm, *Morality, Mortality*, vol. 1 (Oxford: Oxford University Press, 1993), pp. 156-61; *Intricate Ethics* (Oxford: Oxford University Press, 2007), pp. 31-40; Temkin, *Rethinking the Good*, ch. 3; and Alex Voorhoeve, "How Should We Aggregate Competing Claims?," *Ethics* 125 (2014): 64-87. I justify my own theory of limited aggregation that coheres with Scanlon's latest revision of contractualism in Chapter 5 (*Aggregation, Balancing, and Respect for the Claims of Individuals*).

identified victims and statistical victims alike, even though equal respect for identified and statistical victims was one of the key motivations for ex post contractualism. Suppose that in a one versus one confrontation a complaint against missing fifteen minutes of a World Cup match is as strong as a complaint against a risk of death of 1 in 40 million. If we can either save one person from missing part of the match or one person from this risk of death, we should be indifferent. If, however, there were two people subjected to this risk of death, we should save them at the expense of the person missing parts of the World Cup match. Now what if there are many people who would be missing fifteen minutes of the World Cup match? It seems that here numbers should matter. Otherwise we would give undue importance to small risks. We should rather spare a million people of missing the World Cup match, then to reduce a 1 in 40 million risk of death to a single person. In other words, here we should be allowed to aggregate the complaints against missing parts of the World Cup match. If this is so, then we should be allowed to aggregate both the complaints against the risk of death and the complaints against missing fifteen minutes of the World Cup match. If there are many complaints against small risks, similar to my *Gas* case, then these might add up to one expected life lost. But since we are also allowed to aggregate the complaints of the World Cup viewers, these might be decisive. However, if we contrast a single identified person with the World Cup viewers, as in *Jones*, we are required to save the identified person. Distinguishing between different kinds of harm can therefore not treat cases where a statistical life is lost the same as cases where an identified life is lost.

Third, the idea that heavily discounted complaints against serious harm remain morally significant is also implausible in its own right. One downside of this view is that it has a problem analogous to Kamm's Sore Throat Case. In Kamm's original case we have a choice between saving one life and saving another life *and* saving someone from a sore throat. Kamm wants to say that here we should not decide in favor of saving the second person's life solely on the grounds that we can also save someone from a sore throat.¹⁹ Now imagine that the tiebreaker is not the sore throat but the imposition of a tiny risk of death, for example, by calling an

¹⁹ Kamm, *Morality, Mortality*, 1:146-47.

ambulance. Not only is it the case that we are then *permitted* to save the person who does not require the ambulance on grounds that her rescue does not impose a trivial risk. Even further, we are *required* to save her. It would be impermissible *not* to use the trivial risk as the deciding factor. Together with the insufficient motivation for treating equally strong complaints differently, I think this gives us grounds to treat equally strong complaints as either relevant or irrelevant. What we should accept, however, is that complaints can be aggregated when their strength is relevant to the strength of the complaints with which they are competing.

Since the ex post contractualist cannot distinguish between the aggregation in *Gas* and the aggregation in *Jones*, she should accept the risk imposition in *Gas* as permissible. She then cannot accept the principle of *Equal Treatment for Equal Statistical Loss*. This is bad news for the ex post contractualist for two reasons. First, she must reject the plausible claim that harms have the same disvalue regardless of how many other people will also be harmed. The risk that one person will be harmed will receive greater weight than the risk that any additional victim over and above the first victim will be harmed. Second, a version of ex post contractualism that accepts the risk imposition in *Gas* includes a bias against statistical lives, a charge ex post contractualists usually raise against their ex ante colleagues. In some cases, like *Gas*, a statistical life will not be saved even though an identified life would have been. This criticism against the ex ante view becomes less convincing, since the two theories differ only in the degree to which they are biased against statistical lives.

IV. What We Owe ... to Whom?

My discussion of the sequence has revealed two things: First, it has shown that two plausible principles that an ex post view would want to fulfill cannot be jointly fulfilled. Second, it has given us a better way of understanding ex ante and ex post views. We can understand these views as answering the question of whose complaints we should be concerned with as contractualists. Should we appeal to the complaints of identifiable individuals (epistemic ex ante)? Should we appeal to the complaints of rigidly designated individuals (objective ex ante)? Should we appeal to

the complaints of non-rigidly designated individuals (ex post)? In what follows I will argue in favor of objective ex ante contractualism. The concern with the complaints of rigidly designated individuals expresses the best model of acting in ways that are justifiable to each separate person. As I explained earlier, such a concern with rigidly designated individuals means that we should draw a distinction between cases involving epistemic and cases involving objective risk. In a second step, I argue that this is a virtue of objective ex ante contractualism since it illuminates the distinction between luckless and doomed victims.

A. *Justifiability to Each Separate Person*

The core idea of contractualism is that actions must be justifiable to each. Moreover, in order to respect the separateness of persons our actions must be *justifiable* to each as a *separate* person. This guiding idea, I argue, supports the view that our justification should address rigidly designated individuals rather than identifiable individuals or non-rigidly designated individuals. In other words, the basic idea of contractualism supports objective ex ante contractualism.

Consider the difference between the following three statements made by the U.S. President after deciding on which option to take. The three statements mirror the three options for who the ideal addressee of justification is. In each scenario the President addresses a victim and tries to justify the imposition of the burden on her.²⁰

A: "To Charlotte Williams, born on the 1st of June 1975, resident of Santa Barbara, who is going to die from this measure, I can only say that I am deeply sorry but that your complaint against the measure was outweighed by other complaints. Even though it is hard to accept, I am convinced the measure is justifiable to you too."

B: "To the Californian with the genetic predisposition, whoever he or she *may be*, I hope that you hear me. I can only say that I am deeply sorry but that your

²⁰ I grant that this is the least plausible part of my dissertation and stretches the imagination of the reader. I invite the reader to imagine *another* President making these compassionate and carefully crafted words to make it more believable.

complaint against the measure was outweighed by other complaints. Even though it is hard to accept, I am convinced the measure is justifiable to you too.”

C: “To the Californian who is going to die from the measure, whoever he or she *turns out to be*, I can only say that I am deeply sorry but that your complaint against the measure was outweighed by other complaints. Even though it is hard to accept, I am convinced the measure is justifiable to you too.”

Should we believe that there is an important moral difference between justification A and justification B? Epistemic *ex ante* contractualists like Johann Frick believe that there ought to be. Frick, for example, holds that it makes a difference whether or not we can identify a given individual with a complaint. Should it be impossible or overly burdensome to identify which person is going to die from the proposed policy, then we ought to treat this as a case of many discounted complaints against killing.²¹ I disagree. Frick’s argument relies on an idea about what we can justify to each person. But this, I think, misrepresents the core idea of contractualism. Contractualism is about *justifiability* rather than *actual justification*. Justifiability is already an idealized concept. It requires us to take into account all effects of actions on everyone concerned and to take into account all complaints everyone may have. It also requires us to take into account complaints that no one in fact has or will raise. The ideal of justifiability is one of acting in accordance with principles that would sustain a hypothetical and ideal form of justification. Since we have already idealized, it is difficult to see why we should not idealize epistemic limitations as well.

Therefore, I believe that we should think of A and B as equally good justifications. In both cases the President is justifying her behavior to the victim. Both speeches are meant for one person alone and address and justify the action to one person alone. The only difference is that speech A includes more detail that allows us to identify the individual. While identifiability is important for Frick, he does not discuss what is required to identify an individual. Taking a cue from Casper Hare, we can think of “identifying” an individual by knowing more personal information

²¹ Frick, “Contractualism and Social Risk,” pp. 193-94.

about that particular person.²² We might then have identified a victim without knowing their name as long as we know enough distinctive personal information. But whether or not the President is able to include more detail in the description, such as name, birth date, place of residence or other identifying information, is morally irrelevant. We are not interested in token individuals because of names or other personal information such as appearance, tastes, or talents that allow us to identify them. This information is morally superfluous. We are interested in token individuals because of their particular situation and predicament. The description “the Californian with the genetic predisposition” conveys everything that is morally important. Objective ex ante contractualism bases its complaints only on morally relevant information about a person’s situation. This ensures that we do not confuse justifiability which is at the heart of contractualism with actual justification.

Even more so, at times additional information that allows us to identify individuals can even distort our moral reasoning. Imagine a doctor who has to decide on which treatment to administer to two unconscious patients, Deborah and Eric.²³ Out of expediency the doctor has to administer the same treatment for both, even though they have two different diseases, X and Y. On the one hand, the doctor can think of the prospects that Deborah and Eric have. Without any further information the doctor would assign a 50-50 probability that Deborah has either of the two diseases. (And the same for Eric.) The trade-off between the two diseases will then be regarded as an intra-personal trade-off where Deborah’s and Eric’s interests are the same. On the other hand, the doctor could think of the interests of “the patient with disease X” and “the patient with disease Y”. In this way she would regard the trade-off as inter-personal. This way of regarding the case is superior. The doctor knows that she is dealing with an inter-personal trade-off, she knows that the interests of her two patients are not aligned. Doing one act will harm one and benefit the other. The doctor should not deceive herself into thinking that this is a choice without a conflict of interests.

²² Caspar Hare, “Should We Wish Well to All?,” *Philosophical Review* 125 (2016): 451-72, at pp. 467-71.

²³ The case is a variation of one by Anna Mahtani (“The Ex Ante Pareto Principle,” *The Journal of Philosophy* 114 (2017): 303-23, at pp. 310-11.) Mahtani credits Caspar Hare as her inspiration.

Rather than between A and B, we ought to hold that there is an important difference between B and C. While the contrast between A and B has shown the importance of justifiability as opposed to actual justification, the contrast between B and C shows the importance that justifications have to be addressed to separate persons. In statements B (and A), the President addresses and talks to one person alone, while in C the President does not address any specific person. At the time of the President's address, the words are not addressed to one individual alone. The first two speeches constitute a private channel of communication between the President and the victim. The communication and the justification are one-to-one. If what the President says is correct, then she would have succeeded in justifying her action to this person.

In the third speech, however, the words cannot address only one person. The justification cannot be private or one-to-one in the same sense. At best the President will have addressed a person once the policy is applied, but this does not make it the case that the President did address this person prior to the action or when acting.²⁴ It is thus difficult to see how the justification in C conforms to the contractualist ideal of justifying one's action to each. Justification is owed to each separate person. But the discourse in C does not address persons separately. The appeal of a justification like C stems from the way we assimilate this thought with justifications given along

²⁴ The formulation here implies a rejection of the view that future contingents already have truth-values. But my argument is not restricted to this metaphysical view. Some philosophers believe that future contingents already have truth-values and that this view is compatible with indeterminism (see Nuel Belnap and Mitchell Green, "Indeterminism and the Thin Red Line," *Philosophical Perspectives* 8 (1994): 365-88; or David Lewis, *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986), pp. 206-9). If this is true, then it is the case that the President's justification does actually address one individual even though the identity depends on the objectively risky event. However, this only holds *if* the President *actually* acts this way. Should the President decide not to act this way, we have to assess a counterfactual rather than a future contingent. Under most standard views of counterfactuals these counterfactuals will be open counterfactuals without a truth-value (see Caspar Hare, "Obligations to Merely Statistical People," *The Journal of Philosophy* 109 (2012): 378-90, at pp. 380-82). This means that the model of justifiability used in C and whether it addresses a person will depend on what the decision-maker ends up doing. But this puts the cart before the horse. An action should not be more or less justifiable based on what the agent actually does. The fact that alternative actions will be open counterfactuals also means that the model of justification used in C cannot be applied to help decide between different alternatives, since all but one of the alternatives include an open counterfactual.

the lines of my proposed speech B. In these cases the “someone” refers to a given individual. But this is not the case in C. In C, the justification addresses a compound of different individuals across different possible worlds.²⁵

We can see this even more clearly when we consider cases where the complaint of “the Californian who is going to die” outweighs the complaint of a rigidly designated individual, such as Bob Johnson. Bob Johnson could rightly ask who the person is that can reasonably reject the proposal that would get him off the hook. It cannot be that we determine the identity of said person only after the fact. Even more so, ex post contractualism makes it impossible for us to know or determine who that person would be. It would be morally impermissible to perform the actions which uniquely could determine the identity of this person. It will never be determined who the person was for whose sake we sacrificed Bob Johnson’s leg.

Indeed, there is a compelling justification for imposing risks in cases like *Wheel* even though we know one person will be harmed. Note that no individual victim in cases like *Wheel* would have been permitted to save herself over Bob Johnson. She was facing only a small risk of death, a risk small enough that she would have been required to bear this risk. We can give the following powerful reason to the victim: You were not allowed to save yourself even accounting for your partiality towards yourself. So, you cannot complain to a third party that was not allowed to be partial towards you, that she did not save you.²⁶

The fact that speech C, and thereby the model of justifiability ex post contractualism employs, fails to address a particular person can also be seen clearly in a different context. By carrying the logic of speech C forward ex post contractualism makes the permissibility of risk impositions dependent on mere population size. For this see the following case:

Water (County Level). There is a toxic pollutant in the groundwater all over California. The pollutant will lead to every Californian losing the small finger of the right hand if nothing is done. Scientists have developed a chemical that will neutralize the pollutant. However, the chemical is still

²⁵ See also Frick, “Contractualism and Social Risk,” pp. 196.

²⁶ See also Voorhoeve, “How Should We Aggregate Competing Claims?,” p. 74.

in development and thus risky. The scientists have reduced the risk of death considerably to only 1 in 40 million. The risks are objective and probabilistically independent for each Californian. While the pollutant affects the groundwater of all of California, the water systems are separate for each county. Each local authority has to make the decision.

Let us take as an example Santa Barbara County which has only about 450,000 residents in contrast to the 40 million residents of California as a whole. The objective risk for each individual to die is still 1 in 40 million. But while the likelihood of at least one person dying is significant across California, the likelihood of at least one person dying in Santa Barbara County is now lower. The probability is only slightly over 1 percent. Perhaps discounting the harm of death by 99 percent makes the harm less grave than the loss of the finger. (If you do not believe the harm is discounted enough, just reduce the population size further.) If this is the case, then ex post contractualism allows releasing the chemical for Santa Barbara County. If all the other counties are of a similar or smaller size than Santa Barbara, the risk imposition would be permissible there too.²⁷

This leads to an absurd conclusion. Ex post contractualism needs to hold the following. If the government of California were to decide, releasing the chemical would be impermissible in the contractualist sense; it would not be justifiable to each. If each local government were to decide, releasing the chemical would be permissible in each case. It would be justifiable to each. Even though every single person is affected in the very same manner, the policy would turn out to be unjustifiable to one of them if the decision was taken at a different level. Ex post contractualism somehow generates a person with a complaint from a group of persons without a complaint. The absurdity is even clearer if we accept that unjustifiable risk impositions are wronging an individual.²⁸ While none of the county governments would be wronging

²⁷ Some counties of California are comparably large, e.g. Los Angeles County with over 10 million people. We can imagine that in those counties more local authorities have to make the decision.

²⁸ See e.g. John Oberdiek, *Imposing Risk* (Oxford: Oxford University Press, 2017), pp. 126-53. Frances Kamm has argued for the more radical claim that Scanlon's account for wrongness should generally be understood as an account of wronging (Kamm, *Intricate Ethics*, pp. 461-68).

an individual if they released the chemical, the Californian government would be wronging an individual. But who would be wronged? This reveals that ex post contractualism fails to give us a model of acting in ways that are justifiable to *separate* persons.

B. *The Luckless and the Doomed*

Objective ex ante contractualism draws a distinction between cases like *Dust* in which the risk imposition is epistemic and cases like *Wheel* in which the risk imposition is objective. This is because in cases of epistemic risk, like *Dust*, we can identify a rigidly designated individual who is certain to be harmed while in cases of objective risk, like *Wheel*, we cannot. This distinction may seem suspect and none of the other authors writing on contractualism has considered it relevant.²⁹ However, far from being a defect of the view, I believe that distinguishing between epistemically risky cases and objectively risky cases is a virtue of the view. The reason is that the distinction tracks another distinction about the moral relevance of luckless and doomed victims. In epistemically risky cases like *Dust* there is going to be one doomed victim while in objectively risky cases like *Wheel* there is going to be one luckless victim. While the effect on both is the same, we can see that there is a significant difference between having doomed a person who ends up dying and having given that person a very favorable chance of survival.

John Broome in his discussion of fairness makes the following remark about persons who lose out in the allocation of a scarce good.³⁰ Whoever loses out has grounds for complaint. But the person would have an even bigger ground for complaint if it was never even on the cards for her to have received the good. We cannot justify our allocation to this person by saying that we gave her a fair shot at receiving the good. Losing out for this person is not “tough luck” but, worse, an inevitable feature of our decision. The fact that she may have won, that it once was on the cards for her to win, mitigates her complaint against missing out. In short, after

²⁹ Indeed, Frick argues against its relevance in “Contractualism and Social Risk,” pp. 197-201.

³⁰ John Broome, “Fairness,” *Proceedings of the Aristotelian Society* 91 (1991): 87-102, at p. 98.

the allocation a luckless loser has a less strong complaint than someone who has been doomed to lose. The lottery example shows how the kind of risk that is at play in allocating the good matters for the complaints that individuals can raise. In a lottery that employs epistemic risks, it was never on the cards for anyone other than the winner to win. In an objectively risky lottery this is not the case. Every person stood a chance of getting the good. The lottery is fair because it is the “luck of the draw” that decides who gets it.³¹ Objectively risky lotteries are such that we can say to the person that she could have received the good. We designed the lottery such that it could have easily gone the other way and she may have won.³²

These points about fairness in allocating goods are not limited to the allocation of benefits. They should also apply to the allocation of burdens or harms. Common examples to illustrate lottery fairness include such cases. The Draft lottery to select soldiers for the Vietnam war is a paradigm example. The cases I have discussed are similar. In all cases harms are avoidable only at the expense of a moral catastrophe. We have to decide about the allocation of harm. This means that we can say to those who are luckless that they could have avoided the harm whereas those who would have been doomed would not have had any such chance. It is a virtue of objective ex ante contractualism that it can distinguish in this manner between luckless and doomed victims.

While the previous considerations on fairness illustrate the importance of the distinction between luckless and doomed in giving reasons after the risk materializes, there are also reasons to care about the distinction before the action. Consider the

³¹ This idea is even invoked by critics who account for lottery fairness in a different manner. George Sher and Michael Otsuka gives accounts of lottery fairness of merely epistemic lotteries since both doubt that lotteries with objective risks exist. Sher mentions the “luck of the draw” interpretation as the most obvious rationale for lottery fairness which is incomplete because it cannot account for the fairness of lotteries that do not employ objective risks. Otsuka argues that objectively risky lotteries would be fairer than epistemically risky lotteries, if it was possible to run them. George Sher, “What Makes a Lottery Fair?,” *Noûs* 14 (1980): 203-16, at pp. 203-4; and Michael Otsuka, “Determinism and the Value and Fairness of Equal Chances,” (unpublished manuscript).

³² I owe this point to Kai Spiekermann. He explores this idea in connection to lottery fairness and social risk in Kai Spiekermann, “Good Reasons for Losers: Lottery Fairness and Social Risk,” (unpublished manuscript).

following case narrated by Anatol Rapoport.³³ In the Second World War an allied air base in the South Pacific faced the problem that most of their planes did not survive their allocated missions. The chance of survival was only one in four. An alternative but rejected policy would have increased the chances of survival. Only half of the planes would fly missions with increased bomb load. The increased load would mean that less fuel would be available and the pilots could not return to safety and would crash. Instead of giving everyone a chance of one in four, the policy would fate half the pilots to certain death. The repulsion against and failure to adopt the policy is best explained by an objection against dooming individuals to death.³⁴

However, the difference between doomed and luckless victims goes beyond cases where the victims know their fate. Assume a small variation of this case where, in order to ensure compliance, after the selection by lot all pilots are boarding a plane. The commanders in turn do not know which planes are loaded and which carry empty loads. Pilots who fly an empty plane have orders to return to a different base when they realize their empty load at the first target. At the decision to order the pilots to fly, every pilot faces an epistemic risk of death of 50 percent. This variation is no less objectionable than the initial plan. By distinguishing between doomed and luckless victims, objective ex ante contractualism can account for this. The doomed pilots are certain to die whereas under the ordinary protocol all pilots face a three quarter objective risk of death. By contrast, epistemic ex ante contractualism may justify the order to fly given that it reduces the epistemic risk each pilot faces. Ex post contractualism in turn would justify the order to fly given that it reduces the number of expected lives lost. Only objective ex ante contractualism can account for the answer which is both the actual decision at the base and the intuitively correct one.

One might object to my analysis of the case of the pilots. Assuming that the selection by lot is random, every pilot would have faced a 50 percent objective risk of death under the alternative policy as opposed to a 75 percent objective risk of death

³³ Anatol Rapoport, *Strategy and Conscience* (New York: Harper & Row, 1964), pp. 88-90. Rapoport presents this case as a real-life case but could not vouch for its authenticity.

³⁴ Jonathan Glover reports that the horror of certain death motivates the refusal to accept the policy of one-way missions in Rapoport's example. Jonathan Glover, *Causing Death and Saving Lives* (London: Penguin Books, 1977), pp. 212-13.

under the standard policy. However, it is not accurate to draw the conclusion that objective ex ante contractualism would therefore endorse the alternative policy. The problem here is similar to the problem of medical experimentation discussed by Frick. In the example of medical experimentation there is an ex ante selection of persons to be experimented upon. At the stage of selection the policy of experimenting is beneficial to all, but after the selection is made, severe hardship is imposed on some. Objective ex ante contractualism can avail itself to the same reply as epistemic ex ante contractualism and adopt what Frick calls the Decomposition Test.³⁵ The Decomposition Test imposes a requirement to always act, in each action, in ways that are justifiable to each. The policy of selecting people at random first and then imposing severe hardships on them does not meet this test. This holds for the case of medical experimentation as well as for the case of the pilots. When sending out the pilots to fly, some pilots are doomed to certain deaths. Objective ex ante contractualism prohibits this.³⁶

Our objection to dooming the pilots to certain death are linked with our intuitions about risk concentration and risk dispersal. Take, for example, our reaction to a now debunked story about the Coventry Blitz, the horrendous bombing raid of Nazi aircrafts on the city of Coventry. According to the story Churchill knew about the impending devastating attack on Coventry and could have averted it. In order to not reveal military intelligence, Churchill sacrificed Coventry for the sake of the overall war effort and reducing the overall death toll. When the story was published it was perceived as a grave accusation and moral flaw for Churchill to have acted this

³⁵ Frick, "Contractualism and Social Risk," pp. 201-12.

³⁶ Nir Eyal has suggested that what is problematic with Rapoport's case is not that the pilots are doomed, but rather that they are doomed by their commanders. The commanders, as opposed to enemy fire, would be killing the pilots by adopting the policy. See Nir Eyal, "Concentrated Risk, the Coventry Blitz, Chamberlain's Cancer," in Cohen, Daniels, and Eyal, *Identified versus Statistical Lives*, pp. 94-109, at pp. 105-7. However, I believe that this part of the story is not central. My reaction would not change if some of the planes had insufficient fuel due to sabotage and the commanders had the choice of aborting the mission and calling the planes back. (Imagine that bombs are loaded automatically according to overall weight.) The commanders would still doom some pilots to certain death, even if the pilot would not be killed by the commanders.

way.³⁷ Distinguishing between doomed victims in Coventry and unlucky victims elsewhere in the United Kingdom can explain why. Rapoport's pilot case as well as the Coventry Blitz reveal that our intuitions about concentrating and dispersing risks are sensitive to what kind of risk we are talking about. The plan to fly one-way missions disperses and reduces epistemic risks, but this does not make the plan very appealing given that objective risks are concentrated. There is little point in dispersing epistemic risks if we knew that it is already carved in stone who will die. However, dispersing objective risks is a genuine sense in which burdens are shared and additional burdens are spread more widely.

Thus far I argued that part of the reason why the distinction between objective and epistemic risks is meaningful is because it can explain the moral difference between luckless and doomed victims. This allows me to respond to one concern about my view. Imagine a vaccine that we know carries a certain small risk of serious harm. Whether or not the foreseen harms of mass vaccination are a reason against the mass vaccination will depend, on my view, on the specific mechanism by which the risk manifests itself. If the mechanism is a random mutation, then it is a small objective risk whereas if the mechanism relies on genetic predispositions, then it is a small epistemic risk but a large objective risk. Why should this mechanism matter? In response: The mechanism matters because in the case of the random mutation the harmed victim is luckless whereas in the case of the genetic predisposition we would doom the victim to be harmed. As I have argued, there is an important moral difference between luckless and doomed and this moral difference makes the otherwise uninteresting seeming difference in the biological mechanism of the vaccine relevant. While often we do not know with certainty what mechanism applies, we often have information whether our applied case is more like the case of random mutations or more like the case of genetic predispositions. This, I believe, rightly influences how we ought to act in the case.

The distinction between objective and epistemic risks is also important for another reason. It can illuminate the importance of hypothetical consent. An

³⁷ See Eyal, "Concentrated Risk, the Coventry Blitz, Chamberlain's Cancer," pp. 94-95. Eyal seeks to vindicate Churchill's imagined reasoning.

important and familiar reason for rejecting ex post contractualism is that it makes actions impermissible even if these actions would receive the hypothetical consent by all affected parties. For each individual it is sometimes rational to take small risks of death for moderate gains. For example, it would be rational to take a vaccine against a disease that is not life-threatening even if there is a risk of a lethal allergic reaction. If such risks are imposed on a large scale, then we can be virtually certain that some person will die from the risk. Not only are these risk impositions intuitively permissible, but we can give a strong argument in favor of them. Frick has called this the Argument from the Single Person Case.³⁸ If the risk imposition were to affect only a single person, it would be permissible. In such a case it seems reasonable that we should do what is in that person's rational self-interest. Now in a second step, we learn that there is a second person in an identical position from the original person. The risky treatment is available at no additional cost for that person too. The case is still relevantly similar to deciding for one person. It does not involve any competing claims. We can add more and more people. Individually, we would always favor giving them the treatment. Yet ex post contractualism needs to hold that for a sufficiently large group the risk imposition becomes impermissible.

Is there anything the ex post contractualist could say to reject the Argument from the Single Person Case? The best response seems to be the following. The hypothetical consent that each person would give is vitiated because they are imperfectly informed.³⁹ If we knew that a person would only consent because she is insufficiently informed, it is less plausible to assign moral weight to this hypothetical consent. Imagine that you are a guardian charged with that person's interest. If you were fully informed and knew that the risk imposition is in that person's interest only because of imperfect information, you would not assign moral importance to that fact about self-interest. A close variation of this case is a case where you are in charge of various person's interests. You may not know which person is going to lose out, but

³⁸ Frick, "Uncertainty and Justifiability to Each Person," pp. 133-34; and Frick, "Contractualism and Social Risk," pp. 186-88. Similar arguments are made by Tom Dougherty ("Aggregation, Beneficence and Chance," *Journal of Ethics and Social Philosophy* 7 (2013): 1-19, at pp. 8-11) and Caspar Hare ("Should We Wish Well to All?," pp. 455-67).

³⁹ Fleurbaey and Voorhoeve raise this criticism. See Fleurbaey and Voorhoeve, "Decide As You Would With Full Information!".

you still know the related fact that one of the persons whose interests you look after is going to lose out. As a fully informed guardian you would therefore object to the action. In epistemically risky cases like the vaccine case this is the case. Somewhere in the chain there is a person for whom it is not in their fully informed self-interest that the risk will be imposed. The chain of single person cases is no longer fully symmetrical under conditions of full information. Since we can anticipate this already, we have grounds to object to the risk imposition.

The reply to the Argument from the Single Person Case helps us refine the importance of hypothetical consent. Unlike actual consent, we have no reason to give moral significance to hypothetical consent that arises due to imperfect information. Yet this challenge does not impede giving significance to hypothetical consent which is not tainted in this manner. This is the case for objectively risky cases. Remember the *Water* case I introduced earlier. In *Water* every Californian faces the same problem for deliberation. Either they will lose their small finger or they incur a minute risk of death. The risk at stake here may be in the neighborhood of many risks that the Californians voluntarily incur on a regular basis for small benefits. The gamble is in the self-interest of each Californian; each would hypothetically consent. In this case the response that hypothetical consent arises only out of imperfect information has no bite. Even if all Californians knew all relevant facts about themselves, it would nonetheless be in their self-interest to take the gamble. The Argument from the Single Person Case stands. Distinguishing between objective and epistemic risks helps us understand that the Argument from the Single Person Case is compelling in some cases while unconvincing in others. By distinguishing between these cases, objective ex ante contractualism retains what is attractive in the Argument from the Single Person Case while avoiding the charge that hypothetical consent is vitiated due to imperfect information. In the revised case all risk impositions are independent from one another. There is no conflict over the resource that gives everyone a favorable prospect for their lives. Since there is no connection between the risks, there is no reason why it should not be permissible to impose all of them at once. Consequently, objective ex ante allows imposing all risks at once.

V. Objections

I will consider three main lines of objection to my version of ex ante contractualism that discounts objective rather than epistemic risk. The first line of objection stems from the possibility that determinism is true. The second line of objection raises objections to the ex ante Pareto principle. The third line of objection criticizes an identified victim bias in my position.

A. Determinism

My view distinguishes between objective risks and epistemic risks. There is a worry that even if this distinction would be of moral importance, it is irrelevant in the real world. If determinism is true, the worry goes, then there is no such thing as objective risk. There might be actually observed frequencies but no objective risk in a robust sense that could be morally relevant. The view that the truth of determinism implies the absence of objective chances was once taken as the orthodox view in the philosophy of probability. Recently, however, there has emerged a growing literature in the philosophy of probability that argues that objective chance or objective probability is compatible with determinism.⁴⁰

⁴⁰ See Barry Loewer, "Determinism and Chance," *Studies in History and Philosophy of Modern Physics* 32 (2001): 609-20; Carl Hoefer, "The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance," *Mind* 116 (2007): 549-96; Luke Glynn, "Deterministic Chance," *British Journal for the Philosophy of Science* 61 (2010): 51-80; Antony Eagle, "Deterministic Chance," *Noûs* 45 (2011): 269-99; Michael Strevens, "Probability out of Determinism," in *Probabilities in Physics*, ed. Claus Beisbart and Stephan Hartmann (Oxford: Oxford University Press, 2011), pp. 339-64; Nina Emery, "Chance, Possibility, and Explanation," *British Journal for the Philosophy of Science* 66 (2015): 95-120; Roman Frigg and Carl Hoefer, "The Best Humean System for Statistical Mechanics," *Erkenntnis* 80 (2015): 551-74; Christian List and Marcus Pivato, "Emergent Chance," *Philosophical Review* 124 (2015): 119-52. There is a subtle difference in the literature between objective chance and objective probability. Some philosophers have argued that while there might be objective probabilities, these probabilities do not express the true randomness that is associated with chance (see Aidan Lyon, "Deterministic probability: neither chance nor credence," *Synthese* 182 (2011): 413-32; Strevens sets this issue aside without taking a stand, see Strevens, "Probability out of Determinism"). Since these probabilities are nonetheless objective and features of the world, my arguments may still apply to this type of objective risk.

A first reason to think that the objective probabilities are compatible with determinism stems from the existence of probabilistic laws in science. To give some examples, classical statistical mechanics, evolutionary theory, Mendelian genetics, meteorology and the social sciences all include probabilistic laws. In fact, it appears that deterministic laws are largely confined to just one branch of science, namely the physical sciences. The probabilities posited by the laws of the special sciences, including parts of the physical sciences like classical statistical mechanics, do not appear to be epistemic. For example, the process of ice cubes melting when being put in water is a probabilistic process according to classical statistical mechanics. It appears that classical statistical mechanics can, by virtue of this probabilistic law, *explain* why the ice cube is melting. Indeed, if we believe that special sciences above the micro-physical level are able to explain phenomena, then they explain these phenomena by reference to probabilistic laws. This makes it difficult to conceive of such laws as being concerned with epistemic probabilities. The laws of classical statistical mechanics cannot both incorporate our ignorance about deterministic processes and at the same time explain why ice cubes are melting or why the climate system is changing. Our ignorance cannot explain.

So how can we accommodate the fact that laws of the special sciences posit objective chances with the idea that the universe is deterministic at the micro-physical level? One rationale for the compatibility of objective chance and determinism at the micro-physical level is that the descriptions of “chance” and “determinism” are level-specific.⁴¹ It is imprecise to talk about whether or not the world is deterministic. The real question is whether the world is deterministic or not *at a specific level*. A helpful test to see whether or not the world is deterministic at a given level is to ask whether knowing the entire history of the world described at that level determines a future event. Those who argue that the world is deterministic at the micro-physical level mean to say the following: If we knew all the laws of nature as well as the initial conditions of the universe described in micro-physical language, then the only chances of an event happening are zero or one. But this does not say anything about whether or not the world is deterministic at some macro-level. It does not follow that,

⁴¹ Glynn, “Deterministic Chance”; List and Pivato, “Emergent Chance”.

at the macro-level, the history of the world already determines the event. In other words, determinism at the micro-physical level can coexist with indeterminism at some macro-level. This way macro-level events like melting ice cubes or coin tosses will have their own macro-level chances.

For the purposes of moral theorizing, we are predominantly concerned with the agential level, the level at which we describe agents and their actions. The agential level is the appropriate level for the moral decision-making of agents. What would rule out the possibility of objective chances in the relevant sense is, therefore, not determinism at the micro-physical level but rather *determinism at the agential level*. Yet there is no reason to think that our world is deterministic at the agential level. To the contrary, all indications of our best available (social) science at the agential level tell us that the world is *indeterministic* at the agential level. Even if we knew the entire history of the universe described at the level of agents and macro-objects like coins together with all laws of human behavior, we would not be able to predict, say, the outcome of the next Presidential election. Arguments for determinism rely on information about micro-physical particles and their properties, something that is inadmissible when thinking about whether the world is deterministic at a higher level. The level-specific approach to determinism and chance retains the ability to draw a distinction between objective chance and epistemic credence at each level of description.⁴² Imagine an agent is about to toss a fair coin. The odds of the coin landing heads are 0.5. These are objective chances since the prior history of the world, at the level of coin tosses, does not determine this event. After the coin toss the agent is covering the coin and asks again what the odds are of the coin having landed heads. The answer would seem to be 0.5. But this statement about probabilities is clearly different from the earlier one. The second odds are credences, the first are chances. Thus, the level-specific view can retain the distinction between chances and credences at every level. This distinction in turn means that while agents can create objective chances, they can also create merely epistemic risks. A lottery based on whose birthday is earliest in the year would create epistemic risks if the birthdays of participants are unknown, but it would not create objective risks for the participants.

⁴² See List and Pivato, "Emergent Chance," pp. 139-42.

We can see the point of the level-specific view in another way. Consider again the coin flip. Assume that we hold all other factors constant except for the force exerted on the coin. The following conditionals might all be true:

“If I flip the coin with a force between 0.18345 and 0.18348 N, it will land heads.”

“If I flip the coin with a force between 0.18349 and 0.18352 N, it will land tails.”

“If I flip the coin with a force between 0.18353 and 0.18356 N, it will land heads.”

And so on. But what about the conditional “If I flip the coin, it will land heads”? Or the conditional “If I flip the coin, it will land tails”? The antecedents of these conditionals are underspecified. They do not tell us with which force the coin is flipped and the deterministic laws of physics tell us that small changes in the force applied to the coin lead to different outcomes. The antecedents of the underspecified conditionals describe a set of possible worlds. In this set there are some possible worlds where the coin lands heads and some possible worlds where the coin lands tails. What we can give for the underspecified conditional is a probability of how many worlds are head-landing worlds.⁴³ The fact that this probability is not merely epistemic can be seen if we consider the case in which the conditional is a counterfactual conditional. Processes like this coin flip are counterfactually open. No head-landing world is relevantly more similar to our actual world than any tail-landing world. Since the process is counterfactually open, there will not be a fact of the matter about what would have happened had we flipped the coin. There would only be a counterfactual probability. Since there is no fact of the matter what would have happened, this probability cannot be interpreted to refer to our ignorance about what would have happened.

Now why should we be interested in underspecified conditionals as opposed to fully specified conditionals? After all, in a conditional that is specified at the micro-physical level there are no non-trivial probabilities, if we assume determinism at the

⁴³ See also Caspar Hare, “Obligation and Regret When There is No Fact of the Matter About What Would Have Happened if You Had not Done What You Did,” *Noûs* 45 (2011): 190-206, at pp. 190-94; and Hare, “Obligations to Merely Statistical People,” pp. 380-82.

micro-physical level. The reason is the link between contractualism and evidence-based criteria of rightness. Risk impositions are only an issue for contractualism if it is interpreted as an evidence-based criterion of rightness. Interpreted as a fact-based criterion of rightness, a risk imposition would be wrong if and only if it leads to eventual harm. But a fact-based criterion is unhelpful in guiding the choices of agents. Evidence-based criteria, on the other side, link moral permissibility to a choice an agent can make. They capture morality as answering deliberative questions for agents. The actions that contractualism is concerned with are therefore those that are in the choice set of an agent.⁴⁴ As agents, we are unable to choose the option “flip the coin with a force between 0.18345 and 0.18348 N”. This is simply not an option available to us. The option that is available to us is an option at the agential level, namely “flip the coin”. This gives us an argument for specifying conditionals at the agential level. The agential level captures the options that are available, open to the agent whereas a micro-physical level does not.

The argument for the compatibility of lower-level determinism and objective chances has another upshot. A perennial challenge to ex post contractualism is that it prohibits many intuitively permissible forms of risk imposition where small risks are imposed on large populations. It would seem that traffic victims have reason to reject principles that allow higher speed limits. Major construction works would be impermissible to be built because of the risk of harm to workers. Air traffic may be difficult to be justified because it leads to harms to bystanders. The list goes on.⁴⁵ What these divergent risks all have in common is that they appear random in a relevant sense. They contrast with, for example, the risk of a lethal allergic reaction of an individual. Such an individual’s death may have been difficult to prevent, but it is not random in the same sense. The aforementioned examples all appear random because none of these events is determined by the previous history of the world at the agential level. The event “person is killed in car accident” is not already

⁴⁴ T.M. Scanlon, *Moral Dimensions* (Cambridge, MA.: The Belknap Press of Harvard University Press, 2008), pp. 56-62. This also explains how this argument succeeds if we understand contractualism as a decision procedure for risky cases.

⁴⁵ See Alastair Norcross, “Comparing Harms: Headaches and Human Lives,” *Philosophy & Public Affairs* 26 (1997): 135-67, at pp. 159-67; Ashford, “The Demandingness of Scanlon’s Contractualism,” pp. 298-99; James, “Contractualism’s (Not So) Slippery Slope,” pp. 268-72.

determined by the past history of the world. At most a description of the event in micro-physical language is determined. This means that at the agential level, the level which counts, all the familiar examples are objectively risky. Therefore, objective ex ante contractualism can appealingly explain why it is permissible to impose such risks.

B. *Ex Ante Pareto*

Let me turn to the ex ante Pareto principle. Ex ante Pareto says that if one alternative has a higher expected utility than all other alternatives for all individuals concerned, then it ought to be chosen. While the principle has great intuitive appeal, it has recently come under criticism.⁴⁶ Note that my version of ex ante contractualism differs in two relevant respects from ex ante Pareto. First, the Pareto principle only takes well-being into consideration, while the grounds for reasonable rejection need not be restricted to well-being. Importantly, we should think that different ways of conferring benefits or imposing harms are relevantly different even if they lead to the same outcome in terms of well-being.⁴⁷ Second, the ex ante Pareto principle is often associated with epistemic risks. Some putative counterexamples to ex ante Pareto therefore do not apply to my objective version of ex ante contractualism.⁴⁸

⁴⁶ For example, Matthew D. Adler, "The Puzzle of "Ex Ante Efficiency": Does Rational Approvability Have Moral Weight?," *University of Pennsylvania Law Review* 151 (2003): 1255-90, and *Well-Being and Fair Distribution*, pp. 496-518; Fleurbaey and Voorhoeve, "Decide As You Would With Full Information!"; and Mahtani, "The Ex-Ante Pareto Principle".

⁴⁷ Scanlon, for example, mentions generic reasons of fairness as an example of a ground of reasonable rejection that is not based on well-being. Scanlon's discussion of the relation between contractualism and well-being and his rejection of "welfarist contractualism" is also instructive (Scanlon, *What We Owe to Each Other*, pp. 206-18).

⁴⁸ This includes the mammogram case by Fleurbaey and Voorhoeve who argue that we have broadly contractualist reasons to favor preventive screening since it benefits those who would be worse off otherwise (a group that is already determined). Also, Fleurbaey's and Voorhoeve's objection that ex ante Pareto can violate the guidance given by a fully informed decision-maker depends on an epistemic interpretation of the risk. If the risk is objective, then a fully informed decision-maker would not know which outcome will come about. See Fleurbaey and Voorhoeve, "Decide As You Would With Full Information!". Similarly, interesting questions about the incompleteness of ex ante Pareto only arise under an epistemic interpretation (See Mathani, "The Ex-Ante Pareto Principle").

A main source of worry is that ex ante Pareto (and thus, it is held, ex ante contractualism) admits of large inequality ex post. This is seen most clearly in cases where risks are inversely correlated, we can even be certain that this ex post inequality will arise. However, ex ante contractualism has some resources to alleviate this worry. We should first remind ourselves that complaints are not based exclusively on well-being. The manner in which benefits and harms are distributed matters as well. For example, it seems plausible to say that we have a stronger moral complaint against being harmed intentionally than against the same level of harm when imposed as a merely foreseen side effect. In inversely correlated risks it seems plausible that there is another special causal mechanism at play. The gains to the winner are the causal flipside of the losses of the loser. In other words, the winner gains at the expense of the loser. This peculiar way in which gains and losses are intertwined gives rise to an additional moral complaint.⁴⁹

For each of the two individuals involved in the inversely correlated case it is true that they are subject to a 50 percent chance of losing out *by someone gaining at their expense*. That moral complaint can be articulated by either of the two people involved even before the risk is imposed. We do not need to appeal to the eventual outcome distribution to make this complaint. We do not have to talk about the complaint of “the loser” but can simply appeal to the complaint against the imposition of a risk that someone gains at another’s expense. Thus, in contrast to ex post contractualism, my argument does not imply that cases of inversely correlated risk can be seen as equivalent to inter-personal cases involving certainty.

In cases where the gain in expected utility is modest, this could give us decisive reason not to impose the inversely correlated risk. On the other hand, my reasoning cannot support a preference for the non-risky option when the gain in expected utility is sufficiently great. Yet the ex post model of transforming the risky case into a certain outcome distribution would still counsel for the non-risky option in these cases, provided that the secure option has a higher level of utility than the

⁴⁹ I owe this idea to Thomas Rowe. For further defense see Thomas Rowe, “Risk and the Unfairness of Some Being Better Off at the Expense of Others,” *Journal of Ethics and Social Philosophy* 16 (2019): 44-66.

worse outcome of the risky option. But this preference does not seem justified. Both persons gain something from the inversely correlated risks, namely the prospect of a better life. We should give due weight to this consideration.⁵⁰

C. *Identified Victim Bias*

The third objection arises from the discussion concerning identified and statistical lives. Ex ante contractualism generally favors a bias towards identified lives and has received criticism for giving too strong an endorsement to saving identified lives over statistical lives. Whilst this observation is broadly correct, the relationship between my version of ex ante contractualism and the problem of identified and statistical lives is more complex. Objective ex ante contractualism does not place any emphasis on the victim being identified. Rather, what is relevant is whether the victim is already determined. In a case like *Dust*, we do not have a way to identify the victim but, given that we have a rigid designator for the victim, we should favor her.

Indeed, my proposal can at times account for saving a statistical life rather than an identified life. For this, see a simplified version of a case by Caspar Hare.⁵¹ You have two options, either you head North or you head South. If you head North, you will save one person for certain. If you head South, you can flip an indeterministic coin. If it lands heads, you will save another person. If it lands tails, you will save yet another person. The two potential Southern victims can complain that if you head North they will die. You deprived them of a 50 percent chance to live. They can also complain that you would allocate chances to live more unequally if you were to head North. The potential Northern victim can complain that heading South you deprived her of a 100 percent chance to live. The Northern victim cannot

⁵⁰ Ex post contractualists would reverse their opinion once the complaints against the safe option and the worst case scenario are close enough to be aggregated. The risky option will have more aggregate well-being and presumably be preferred in spite of the complaint against gaining at the expense of someone. However, it seems more plausible that our judgment should be reversed to favor a risky option not because of the aggregate well-being but rather because both individuals receive a valuable chance of a better life.

⁵¹ Hare, "Obligations to Merely Statistical People," pp. 382, 385.

raise an additional complaint about the unfairness of the unequal distribution of chances. If we accept limited aggregation, then it seems plausible that a complaint against a 50 percent chance of death is close enough to a complaint against a 100 percent chance of death. If this is correct, and we are permitted to aggregate the claims of the Southern victims, then the added complaints against unfairness would tip the balance. It would follow, on my view, that you ought to head South and save the statistical rather than the identified life.

Nevertheless, the general observation is correct. Ex ante contractualism retains a bias against statistical lives even though this bias is substantially weakened due to the permissibility of limited aggregation. Take, for example, the following revision of *Wheel*: The indeterministic roulette wheel does not release one ball but ten balls that will kill ten different persons. To many it is difficult to accept that we should prioritize Bob Johnson's leg over multiple statistical victims. However, we should note that the individual risk for each person, while higher than in the standard version of *Wheel*, is still vanishingly low at 1 in 4 million.

On reflection we notice that small risks of serious harms are omnipresent. It is inevitable that large-scale policies will lead to serious harms. In many such cases of social risk, we nonetheless believe that the risk imposition is permissible. Indeed, accounting for these cases is a key challenge to ex post contractualism. Take, for example, the following stylized case:

Vaccine. In order to protect the entire population of California from an infectious disease, which everyone would come down with in the absence of any intervention, the Government is considering a mass vaccination program. The disease is not life threatening but would cause the Californians to limp for two months, similar to the effects of a sprained ankle. While the temporary limp is much less bad than the impairment due to loss of a leg, it is significant enough that the Californians want to avoid it. In extraordinary circumstances, the vaccine can, however, be lethal, although the chance of death for each Californian is only 1 in 4 million. The Government is able to administer the vaccine without intrusion on the bodies of any Californian.

Even though the policy in *Vaccine* will also lead to ten expected statistical deaths, we want to account for the permissibility of *Vaccine*. The risk of death is sufficiently small that it is outweighed by the benefit of avoiding the temporary limp. For example, according to the *National Safety Council*, the odds of a U.S. resident being struck by lightning in their lifetime is a bit over 1 in 180,000, more than 22 times more likely than the harm due to the vaccine.⁵² Rejecting risks of the kind involved in *Vaccine* would make it difficult to pursue many large-scale policies or practices. The challenge is now the following. In the case of *Vaccine*, we prefer saving the population of California from the temporary limp over the loss of ten statistical lives. In the revised *Wheel* case, we prefer saving the ten statistical lives over Bob Johnson's loss of a limb. Now what if we could choose between saving the population of California from the temporary limp or Bob Johnson from the loss of a leg? Since the temporary limp is much less bad than the permanent loss of a leg, it is plausible that a contractualist would reject the aggregation of the complaints against the temporary limp. Hence, we should save Bob Johnson. This leads us to a preference cycle over the three options.

It is not clear how we could justify such a preference cycle. One attempt would be to point out that in *Vaccine* the gamble is in the ex ante interest of all, whereas this is not the case in the revised *Wheel* case.⁵³ This may explain why the option of "ten statistical victims when it was in their ex ante interest to take the risk" is not the same option as "ten statistical victims". I am not convinced that this explains our intuitions well. While it is true that the gamble is in the ex ante interest of all in the stylized *Vaccine* case, I do not believe that this is necessary to the case. I believe that delivering the vaccine would be permissible even if some small and unidentifiable part of the population was already known to be immunoresistant. The vaccine would, therefore, be neither to the ex ante nor the ex post benefit of any of them. In fact, it appears that in most cases of intuitively permissible large-scale risks the benefits are widespread but not universal.

⁵² See the overview at: <https://injuryfacts.nsc.org/all-injuries/preventable-death-overview/odds-of-dying/>.

⁵³ See Alec Walen, "Risks and Weak Aggregation: Why Different Models of Risk Suit Different Types of Cases," *Ethics* (forthcoming).

What the response shows, however, is that it is a mistake to frame the problem in the revised *Wheel* case as either saving ten people from death or one person from the loss of a leg. Such a framing already assumes that what matters is the harm that is the result of the risk imposition. In other words, this framing already assumes the ex post perspective. If my arguments against the ex post perspective are successful, then we should rather phrase this choice as saving the leg of one and reducing the risks of very many by a small amount. So understood, it is more plausible to maintain that it is permissible to impose the risk in the revised *Wheel* case.

We can give the following justification for our choice. At the time of our decision, there was no person who had as strong of a complaint as Bob Johnson did. We were able to justify our action to each of the 40 million persons involved, each of whom faced only a very small risk of death. In fact, none of the 40 million would have been permitted to save themselves from such a small risk if doing so had required the loss of Bob Johnson's leg. For example, each would have been required to call an ambulance to save Bob Johnson's leg even if this would have created a 1 in 4 million chance of being killed by an ambulance sliding out of control. We can acknowledge that a better outcome could have been brought about, in which only one person loses a limb rather than ten people losing a life. But that is the sort of thing non-consequentialists are already willing to acknowledge across a range of familiar cases. Non-consequentialists accept that oftentimes it is impermissible to do what brings about the best outcome because doing so would violate the claims of a single individual. We can understand deontological constraints in this way.

In line with the analogy to deontological constraints, we can accept a further claim. While non-consequentialists accept some inefficiency in terms of failing to bring about the best outcome, non-consequentialists typically accept that there are some cases in which deontological constraints can be overridden. Most non-consequentialists believe that rights may permissibly be violated in cases where doing so is necessary to avoid a moral catastrophe or some other high threshold of weighty moral considerations. In those cases, even deontological constraints such as those which stand in the way of being harmfully used as a mere means can be

exceptionally suspended.⁵⁴ In such cases it can be permissible to do what otherwise would be unjustifiable to the rightsholder, for example, violating the right not to be harmed as mere means. If it is plausible that we can override the individual complaint not to be used as a mere means, then it also seems plausible that we can sometimes override the individual complaint of a determined victim not to be saved. If anything, the complaint against being used as a mere means appears to be a stronger complaint than the complaint against failing to be saved in the cases under discussion in this chapter.

The analogy is strengthened by a deep theoretical connection that contractualism has with a rights-based morality. Contractualism only covers a part of morality, the part that Scanlon identifies with “what we owe to each other”. This part is a part that is concerned with our relations to other persons. A natural thought is when we act in ways that are not justifiable to a given person, we thereby wrong this person. Similarly, when we violate the right of a person, we thereby wrong this person. This suggests an important theoretical connection between contractualism and a rights-based morality given that both are concerned with wrongs done to other persons.⁵⁵ Therefore, the idea that there is some threshold of statistical victims at which point we need to depart from contractualist morality is no more problematic than the widely accepted idea that there is some threshold of bad consequences at which point we need to depart from deontological constraints.

VI. Conclusion

In this chapter, I have argued for a new version of ex ante contractualism which focuses on the complaints that rigidly designated individuals can bring forward. Their complaints ought to be discounted by the objective probability that the harm will come about. Unlike other ex ante contractualists, I do not believe that we should always discount epistemic risk, nor do I believe that we should be

⁵⁴ See e.g. Nagel, *Mortal Questions*, ch. 5; and Judith Jarvis Thomson, *The Realm of Rights* (Cambridge, MA.: Harvard University Press, 1990), ch. 6.

⁵⁵ See e.g. Kamm, *Intricate Ethics*, pp. 461-68.

concerned only with individuals that we can identify. Such an objective version of ex ante contractualism provides us with a plausible model of justifiability to each. It insists that our actions must be justifiable to everyone at the time that we act. It also insists that justification is owed to separate persons. But it does not require the use of morally superfluous, identifying information that would make actual justification to each possible. Objective ex ante contractualism is alone in drawing a distinction between cases in which objective risks are at stake and cases in which merely epistemic risks are at stake. But far from being a defect, this is a virtue. We can thereby illuminate the morally relevant difference between luckless and doomed victims. For these reasons, I conclude that objective ex ante contractualism is a viable and better alternative, which is theoretically superior to both epistemic ex ante and ex post contractualism.

Chapter 4.

Skepticism about Aggregation and Uncertain Rescues

Consider the following (*Anne's Rescue*): Anne is a miner who is trapped in a mineshaft. We can launch a rescue mission that will, with certainty, bring Anne to daylight. If we fail to launch the rescue mission, then Anne will surely die in the mineshaft. However, undertaking the rescue mission has an opportunity cost. Instead of paying for the rescue mission we could use the resources to cure the sore throats of a very large number of people. What should we do? To many it seems that we should save Anne's life. The sore throats are not the right kind of consideration that can outweigh what is at stake for Anne. Regardless of how many sore throats we can cure, we should always save a single life over sore throats. The sore throats do not add up to anything that is of greater moral significance than Anne's life.

However, few actual cases are of this sort. In *Anne's Rescue* we know with certainty what will happen if we launch the rescue mission and what will happen if we provide the pain relief. In the real world, we very often face situations where we are unsure about the results of our action. How should we think about cases like *Anne's Rescue* in circumstances of uncertainty?

The sentiment that sore throats do not add up to the moral significance of a single life expresses skepticism about the permissibility of aggregating harms across different individuals. As I have explained in the introduction to the dissertation, such skepticism is best understood as grounded in the separateness of persons. Skepticism about aggregation can take different forms. A more radical form, which I call *no aggregation*, holds that we should engage in pairwise comparisons between different individuals and never save a person who has a less strong claim to our aid.¹ According to "no aggregation", we should, for example, not save a very large

¹ No aggregation is easily confused with "numbers skepticism", the view that we have no duty to save the greater number. However, it is possible to justify a duty to save the greater number without aggregation, for example, by adopting a Leximin decision procedure. For a discussion on non-aggregative arguments to save the greater number see Otsuka, "Saving Lives, Moral Theory, and the Claims of Individuals," esp. pp. 118-26. No aggregation is a broad tent. Some proponents of no aggregation believe that while aggregating claims is not

number of people from paraplegia over a single person from death. Paraplegia, we can assume, is a substantial harm even if it is much less bad than death for the individual. A less radical form of aggregation skepticism, which I call *limited aggregation*, holds that while it is permissible that the numbers count in deciding whom to save in some trade-offs, in other trade-offs the relative numbers should not count.² For example, the numbers can count only in trade-offs between harms that are relevant to one another. Limited aggregation can then hold that the numbers count in the trade-off between life and paraplegia, but that the numbers do not count in the trade-off between life and sore throats. In line with the acceptance of limited aggregation in the previous chapter, I will focus on limited aggregation.

How should limited aggregation be extended to cases in which we are uncertain about what will happen? This is the question I want to address in this chapter. One idea is that we discount the harms each individual might suffer by their improbability. Anne's claim to aid would then be determined not by the harm she is certain to suffer, but rather by the prospect of harm that she faces. This approach can be called *ex ante limited aggregation*. My previous chapter has defended a contractualist version of *ex ante limited aggregation*.³ A competing approach is the *ex post* approach which determines claims by actual harms and not by prospects. In the previous chapter I discussed *ex post* contractualism. I argued that the most plausible version of *ex post* contractualism embraces a principle that I called *Ex Post Discounting*. The key to *Ex Post Discounting* is that complaints will be assigned to a non-rigidly designated individual such as "the worst-off". But *ex post* limited aggregation need not embrace a contractualist morality where we need to assign complaints to particular individuals. In Section I of this chapter, I begin by outlining

required, aggregative considerations are an intelligible reason that an agent may act upon. See Munoz-Dardé, "The Distribution of Numbers and the Comprehensiveness of Reason". Other proponents of no aggregation believe that in cases of equally strong claims we should give each person an equal chance. See Taurek, "Should the Numbers Count?".

² Kamm, *Morality, Mortality*, 1:156-61; *Intricate Ethics*, pp. 31-77; Scanlon, *What We Owe to Each Other*, pp. 238-41; David Lefkowitz, "On the Concept of a Morally Relevant Harm," *Utilitas* 20 (2008): 409-23; and Voorhoeve, "How Should We Aggregate Competing Claims?".

³ This is in line with most of the discussion on *ex ante* views that are skeptical of aggregation. See in particular James, "Contractualism's (Not So) Slippery Slope"; John, "Risk, Contractualism, and Rose's 'Prevention Paradox'"; Kumar, "Risking and Wronging"; Frick, "Contractualism and Social Risk".

an alternative approach to ex post limited aggregation. I introduce the notion of *ex post claims* and show how an appeal to ex post claims is grounded in the reasons critics of the ex ante approach give for rejecting the ex ante approach. However, building a theory of limited aggregation based on ex post claims leads to a dilemma. I explain both horns of the dilemma in Sections II and III.

I. Ex Post Claims

Ex post limited aggregation is puzzling in one respect. It is a theory about how to decide in circumstances of uncertainty. However, it aims to focus on actual harms as opposed to individual prospects of harm. How is this possible? It may help to consider an example.

Consider the following argument by Marc Fleurbaey and Alex Voorhoeve.⁴ Fleurbaey and Voorhoeve analyze cases such as the following where one of two treatments can be given to two children, Adam and Bill, facing total blindness in the absence of any treatment. With the *egalitarian treatment* both are guaranteed the benefit of having merely a significant, but partial visual impairment instead of full blindness. From a moderate distance they would be unable to recognize a friend but would be able to make out basic shapes. From close distance they would be able to read newspapers, albeit with great difficulty. With the *risky treatment*, we know that one child will end up with good sight and the other child will end up with a visual impairment even worse than the other intervention would have guaranteed him. This unlucky child will only be able to make out basic shapes from close distance. But we do not know which child would be the lucky one.

Fleurbaey and Voorhoeve then reason that there are only two possibilities. Either Adam is the child for whom the risky treatment would be beneficial relative to the egalitarian treatment, or Bill is. If it is Adam, then we know that Bill has a strong claim to the egalitarian treatment. Bill's claim is, we can suppose, stronger than Adam's claim to the risky treatment which would be beneficial to him. If it is Bill, then we know that Adam has a strong claim to the egalitarian treatment which,

⁴ Fleurbaey and Voorhoeve, "Decide As You Would with Full Information!".

once again, is stronger than Bill's claim to the risky treatment. Whichever way things are, or turn out, we know that there are strong claims to the egalitarian treatment.

The claims in Fleurbaey's and Voorhoeve's argument are *ex post* claims. These claims compare how an individual fares given one course of action with how the individual fares given the alternative course of action. Importantly, the *ex post* claim is tied to one state of the world, i.e. tied to one way things may turn out to be. Adam's *ex post* claim is on the assumption that Bill is the one who would benefit from the risky intervention.

The case of Adam and Bill is not the kind of case which invites skepticism about aggregation. There are only two people involved and their respective claims seem relevant to one another. But the case of Adam and Bill illustrates how we can think about *ex post* claims. This idea can be transferred to cases which raise doubts about aggregation. For example, we can imagine that there is a third option on the table, namely, to save neither Adam nor Bill but rather to provide pain relief for sore throats to a large number of people. An aggregation skeptic convinced of the *ex post* approach could reply that we should not choose this option because in either of the two possible states of the world there is a much stronger *ex post* claim advocating for the egalitarian treatment. The claims to sore throat relief would be irrelevant in either state of the world.

While in this imagined variation of Adam's and Bill's case the *ex post* claims unanimously favor one course of action, we can easily imagine cases where this is not the case. What if, for example, there is the third possibility that treatments will be highly inefficient and provide no greater improvement in sight for either Adam or Bill than sore throat relief would give to the others? While in the first two possible states of the world, the *ex post* claims favored the egalitarian treatment, in this third possible state of the world the *ex post* claims favor the sore throat relief. Aggregation skeptics need a theory for how to decide cases like these.

A follower of "no aggregation" could, for example, engage in a pairwise comparison between the strongest *ex post* claim in each state of the world. We would compare, for example, Adam's *ex post* claim to the egalitarian treatment in

S_1 with Bill's ex post claim to the egalitarian treatment in S_2 with the ex post claim of a person benefiting from sore throat pain relief in S_3 . A version of this idea would first discount each ex post claim by the likelihood that their state of the world is the actual state of the world. In either case, one concern with this response is that it takes the idea of pairwise comparisons too far. It makes the decision whom to save dependent entirely on what happens in one state of the world. While using pairwise comparisons in cases of certainty is motivated by respecting the different standpoints of individuals, pairwise comparisons between ex post claims are rather indicative of avoiding a worst-case scenario. In cases in which the worst-case scenario is much worse than all other outcomes, we disregard all the other possibilities and pay attention only to the worst-case. While it may make sense to believe that we should be guided only by the fate of a single individual who has much at stake, it makes little sense to believe that we should be guided only by one possible eventuality.

A more plausible proposal is to adopt limited aggregation and to aggregate ex post claims according to one's favored theory of aggregation. We then need a principle that tells us which ex post claims we can aggregate. There are two possibilities here. First, only the ex post claims within one state of the world can determine whether we can aggregate claims. Second, both ex post claims within one state of the world and across different states of the world determine whether we can aggregate claims. As I shall argue, neither of the options is plausible. This dilemma reinforces the conclusion of my previous chapter that we should reject ex post views in favour of a suitably constructed (objective) ex ante view.

II. First Option: Relevance Tied to a State of the World

According to the first option, whether or not claims can be aggregated is determined solely by reference to the claims in that state of the world. A basic version of this view would tell us to determine first which claims are relevant to the strongest claims in each state of the world, second discount these relevant claims,

and third aggregate all discounted relevant claims. We should then perform the action that satisfies the greatest aggregate of discounted relevant ex post claims.

The proposed view is a natural extension of Alex Voorhoeve's Aggregate Relevant Claims view.⁵ Voorhoeve's view is developed only for cases of certainty. The proposed view supplements Voorhoeve's view with an emphasis on ex post claims and the idea that the relevance of claims is determined only within the same state of the world. The proposed view is also a simplified version of Seth Lazar's Ex Post Maximize Satisfaction of Claims.⁶ Lazar's view is intended to provide a version of ex post limited aggregation.

Consider now the following case (*Uncertain Rescue*): As in *Anne's Rescue* we have Anne, the miner, who is trapped in a mineshaft. Again, we know that the rescue mission will certainly bring Anne's body to daylight and that the opportunity cost is not being able to provide pain relief for sore throats to a very large number of people. However, *Uncertain Rescue* differs from *Anne's Rescue* insofar as there is a very small chance that all help will come too late. Anne might already be dead. Although we heard life signs from Anne only a few seconds ago, it is possible that Anne will have died by the time we reach her. If this is so, there is nothing we can do for her and the rescue mission serves no purpose.⁷ Should this very small chance make much of a difference? It is hard to see why. It is overwhelmingly likely that we can still save Anne and the gains we can achieve by not trying to save Anne are of much less significance than what is at stake for Anne. Importantly, proponents of limited aggregation consider their theory to be of practical relevance. In any real-world scenario there will always be a small chance that rescue will be futile. If limited aggregation fails to account for *Uncertain Rescue*, then it appears to be practically inert.

⁵ Voorhoeve, "How Should We Aggregate Competing Claims?".

⁶ Seth Lazar, "Limited Aggregation and Risk," *Philosophy & Public Affairs* 46 (2018): 117-59, at pp. 139-42. I should note that Lazar ultimately rejects the view but calls it a "real contender" (p. 141). Instead, Lazar embraces a hybrid view (pp. 149-58). This hybrid view contains the ex post component in it. If the ex post view is implausible *as an ex post view*, then this sheds doubt on the plausibility of Lazar's hybrid view.

⁷ We can assume that Anne lives in a society which attaches no special meaning to burial rites. This explains why, if Anne is dead, the rescue mission would have provided not even small benefits to Anne's loved ones in terms of coming to terms with their loss.

Devastatingly, the present option to extend limited aggregation to uncertainty fails to account for this judgment. Here is why. The first step is to identify which are the different ex post claims in each state of the world and which claims are relevant. In S_1 , the state of the world where Anne is still alive, Anne has a strong claim to the rescue option. Her life is at stake. Everyone in the large group has only a small claim in the dispersal of the small benefit. In S_2 , the state of the world where Anne is already dead, Anne has no claim. There is nothing we could do for her. Here as well, everyone in the large group has a small claim in the dispersal of the small benefit. Now are these claims also relevant? In S_2 they are: Since Anne does not have a claim in S_2 , the claims by everyone in the large group are unopposed and thereby relevant. In S_1 there is a competition of claims. Given that Anne's claim is much stronger than the claim of everyone in the group, her claim is the only relevant one.

Anne's claim in S_1 is very weighty, while the claims of the members of the large group in S_2 will carry only little weight. In the second step, we have to discount everyone's claim by the likelihood that their associated state of the world obtains. Anne's claim is discounted by the likelihood of Anne being already dead which is very low. Her claim remains very strong. Discounting the claims of the many group members will further weaken them since the probability of Anne being still alive is very high. Nonetheless, the third step allows us to aggregate all relevant claims in the end. If there are enough members of the group, then they together will outweigh Anne's claim.⁸

⁸ I mentioned earlier that this is a simplified version of Lazar's view. For interested readers, here is how the view is simplified and why this does not affect my argument. (1) Instead of talking about claims and relevant claims, Lazar talks about interests and claims. Lazar believes that we can aggregate all claims, not only relevant ones, but that only some interests are protected by a claim. (2) An interest is protected by a claim if and only if the person whose interest it is would be permitted to save themselves rather than everyone with a relevant competing interest combined. (3) An interest is relevant in turn if that person would be permitted to save themselves rather than the initial person. (4) The differences in Lazar's approach do not affect my argument. In the above argument one can replace "claim" with "interest" and "relevant claim" with "claim" and we have translated the argument into Lazar's approach. The interests of the members of the large group would still be protected by claims because they are unopposed. Anne's interest would similarly survive the more complicated test of Lazar's for being protected by a claim. (5) One further difference is the following: Lazar determines ex post interests counterfactually by comparing the well-being

This does not correspond to what we intuitively thought about the case. The introduction of even a small chance that Anne cannot be helped tips the balance against Anne. Anne is almost certain to be saved from death if we intervene, but the tiny possibility that she may not be makes all the difference here. The problem is that this method allows relevant claims to arise in a given state of the world too easily. Even a fairly small gain can become a relevant claim if it is sufficiently larger than its competitors in that state alone. Provided that there are sufficiently many of these small gains, they can then, in the end, outweigh the relevant claims of other states of the world.

III. Second Option: Relevant Inside and Across States of the World

To avoid the problem that claims can easily arise in one state of the world and outweigh claims in other states of the world, we can opt for the second option that I distinguished. Following this option, it is the interests within the same state of the world and across different states of the world that determine whether aggregation is permissible. This option is not an ad hoc adjustment of our view to avoid the problem I just outlined. It can also be justified by appealing to a core idea of limited aggregation.

In her justification of a limited aggregationist view, Frances Kamm introduces the idea of irrelevant utilities.⁹ The idea is that certain utilities or claims are not important in the face of other more significant claims. To take Kamm's example, it would be inappropriate and disrespectful to consider the claim to being cured from a sore throat when deciding between whom to save from death. A similar idea can be applied to risky cases. It seems inappropriate and disrespectful

of the person given the chosen action with the counterfactual well-being given the alternative action. If we compare the action of providing pain relief for sore throats, we cannot observe, however, whether Anne is dead or alive. So how should we assess Anne's interest here? For Lazar we should take the expected value for Anne, whereas I advocate distinguishing between different states of the world, even if they are epistemically indistinguishable. Given the near certainty of Anne still being alive, this small difference has no bearing on my argument either.

⁹ Kamm, *Morality, Mortality*, 1:144-63.

to consider the claim to be cured of a sore throat in one possible outcome when the other possible outcome is a life and death decision.

Kamm's view has one relevance test in the case of certainty. Claims are relevant, and thus allowed to be aggregated, only if it would not be disrespectful to consider the weaker claim in light of the gravity of the stronger claim. In uncertain cases we could use a two-stage relevance test. Claims have to be relevant to the strongest competing claim within their state of the world *and* across states of the world.

The two-stage relevance test for claims fails. Consider *Desperate Rescue*: We are again uncertain about whether or not Anne, the miner, is still alive in the mineshaft. We have not heard life signs for a long time and the rescue team is losing hope. There is only a very small chance that Anne is still alive, saving her now would be a miracle. The rescue mission is costly, and the recourses could be used to save a large group of people from moderate chronic pain. Moderate chronic pain, we can suppose, is not relevant to death in cases of certainty. Anne's *ex post* claim in the state of the world where she is still alive is the only relevant one. In case that Anne is already dead, the group members have claims to be relieved of the chronic pain. However, none of their claims are relevant to Anne's claim in the eventuality that Anne is still alive. We should try saving Anne, regardless of how likely it is that our intervention will be successful. A very small chance here would make all the difference. The idea that the mere possibility of death should make it disrespectful to consider lesser claims is not convincing either.

A more plausible relevance test is one where the inter-state-of-the-world relevance is determined only after discounting the claims by their likelihood. The idea that considering small claims in the presence of a substantial claim is disrespectful is certainly more plausible when the claims were discounted by their likelihood. This revision also explains what is wrong with the answer that the previous proposal gave in *Desperate Rescue*. Anne's overall claim given the small likelihood that she is still alive is not weighty enough to render the claims against chronic pain relief irrelevant. We can revise the test to a three-stage relevance test. In the first stage we determine which claims are relevant in their state of the world.

In the second stage we discount these claims by the likelihood of their state of the world being actual. In the third stage we determine which claims are relevant to the strongest claim.

The three-stage relevance test struggles to account for cases where risk is dispersed among various states of the world. Consider *Anonymous Rescue*. A large group of people is trapped on a sinking ship. We are able to communicate with the ship and know that at most one person is still alive. We do not know who among the 10,000 crew and passengers is the person who might still be alive. There is also an about 50 percent chance that none of the 10,000 is still alive. We have the choice between a rescue mission or providing a small and certain benefit to a very large group of people. In this scenario there are 10,001 states of the world. In each of the 10,000 states of the world where one person is still alive, that person's ex post claim is relevant and outweighs all other claims. In $S_{10,001}$, the state of the world where there is nothing we can do for the people on the ship, the claims of the large group members are relevant. The ex post claims of each passenger must be discounted by 1 in 20,000. It is quite likely that the ex post claims will then not be relevant to the ex post claims of the group members receiving a small benefit. By dispersing the risk across states of the world, we decide not to try to save the person on the ship. However, if there had been a single, identified person on the ship, her claim would not have been discounted heavily enough to be rendered irrelevant. On the contrary, her claim would have rendered the claims to the moderate benefit irrelevant. Such an identified victim bias is a motivation for ex post views and should not be a component of them.

One way to resist this implication is to protest that my way of setting up the problem was erroneous. It was false to distinguish between the first 10,000 states of the world. A state of the world is a set of possible worlds, or a model of possible worlds, that leaves no relevant aspect of the world undescribed. A state of the world is not a full description of a possible world. One might protest that I *overdescribed* the states of the world. If we should treat the expectation of a 50 percent chance of saving someone to be equivalent to a 50 percent chance of saving a particular person, then this is because the identity of the person to be saved does

not matter. If the identity of the person does not matter, the protest goes, it is because the identity of the person is not a relevant feature in this case. Consequently, the states of the world do not differ in any relevant aspect. What should matter is that some person on the ship might die rather than who exactly has the claim to be rescued. Anonymizing for the victim, the different possible worlds do not differ in any relevant respect.

When we frame the decision problem for risky cases we inevitably have to group possibilities together. Often there will be small differences in possible outcomes that do not have any moral relevance. Our criterion for how to group possibilities will depend on what we think is morally relevant in this case. The fact that in one outcome a shirt will be blue and in another it will be red should not lead us to consider these two as distinct outcomes. The *ex post* proponent can now argue that since we should not be biased in favor of identified lives, the specific identity of a victim does not matter morally either. Hence, we should not divide outcomes where only the identity of the victim differs in different states of the world.

Even though this way of framing the decision problem helps us with *Anonymous Rescue*, it does not help with a related case. In *Anonymous Rescue* all 10,000 people faced the same fate, death. This is why the alternative way of framing the decision problem would only speak of two states of the world, one where someone is alive and another one where no one is alive, both of which are equally likely. Framed this way, we should try to rescue the person rather than giving the small benefit to any number of persons. But plausibly we should also try to rescue one person from, for example, the loss of a limb, rather than giving the small benefit to any number of persons. So, if all 10,000 people are facing the loss of a limb, we can again re-describe this as one state of the world where someone is facing the loss of a limb. Suppose, however, that one of the 10,000 is facing the loss of a limb, another person is facing permanent paraplegia, a third person is facing chronic pain worse than paraplegia, and so on. All 10,000 persons are facing a different harm that is different in morally relevant respects. All 10,000 persons are facing a harm between the loss of a limb and death. In this variation the re-description strategy is no longer possible. These are genuinely different states of the world. The problem of

Anonymous Rescue reappears here. Even worse, if every person were to face the loss of a limb we should try preventing this loss. But if some people are facing a more serious loss, then we should no longer try preventing this loss. If all were to face the loss of a limb, we might be able to reframe the decision problem as having only two states of the world in which case the ex post claim against the loss of a limb is not heavily discounted. But because some people are facing a more severe hardship, the strategy of reframing the decision problem no longer works. Since these people might face a more serious hardship their ex post claims have to be counted as belonging to separate states of the world and discounted separately. This way they become irrelevant.

IV. Ex Post Limited Aggregation Without Ex Post Claims

Thinking about ex post claims leaves us in a dead end. None of the principles that tell us when claims can be aggregated are plausible. Determining when claims can be aggregated only by looking at one state of the world allows aggregation too easily. Determining when claims can be aggregated by looking also at other states of the world makes aggregation either too difficult or too dependent on how the risk is distributed across different states of the world. A common feature of the failure of both approaches is that both give special emphasis to uncertainty. Both treat near-certainty radically different from certainty. The first option radically changed its verdicts once we introduced the small probability of all help coming too late. The second option radically changed its verdict once we introduce the small probability of being able to help at all.

What does this mean for ex post views? One possibility is that ex post views are restricted in their scope. I started my explanation of ex post claims by referring to an argument that Fleurbaey and Voorhoeve bring forward. Fleurbaey and Voorhoeve ultimately argue for what they call the Principle of Full Information.¹⁰ The Principle of Full Information includes a dominance condition. If in all states of the world the ex post claims weakly prefer one action and in no state of the world

¹⁰ Fleurbaey and Voorhoeve, "Decide As You Would with Full Information!," pp. 120-22.

the ex post claims disprefer this option, then we ought to perform the action. If the ex post claims in all states of the world are indifferent, then we ought to be indifferent. Because of its dominance reasoning, the Principle of Full Information does not tell us what to do when different options are preferred by the ex post claims in different states of the world. As I made clear the Principle of Full Information would therefore be silent on all the cases I discussed in this chapter. It would be striking if ex post reasoning would not apply to any of these. Furthermore, it seems concerning for the Principle of Full Information that we cannot expand its core reasoning, the idea of ex post claims, in a plausible manner.

The other alternative is to return to *Ex Post Discounting*. *Ex Post Discounting* is able to resolve the problems of all the cases mentioned here. However, this brings us back to the problems outlined in the previous chapter with ex post contractualism. This provides a challenge for critics of the ex ante view. Critics of the ex ante view would need to show that their objections can provide a foundation for an alternative position. They have, so far, not been able to do this. This strengthens my argument that we should reject the ex post view in favor of a suitable ex ante view. While I explained how objective ex ante contractualism deals with questions of risk and mentioned that it should embrace a form of limited aggregation, I have not yet shown how limited aggregation can be justified. This is the task I take up in the next chapter.

Chapter 5.

Aggregation, Balancing, and Respect for the Claims of Individuals

I. Introduction

One theme in my thesis has been the opposition to the aggregation of harms across different individuals. Such strong resistance to the aggregation of harms across different individuals is the hallmark of a particular kind of non-consequentialism that is inspired by the separateness of persons.¹ Such non-consequentialists object in particular to aggregation when trivial harms might thereby outweigh significant harms.

Few skeptics regarding aggregation believe, however, that we can avoid all forms of aggregation in all cases. Most of these skeptics regarding aggregation would, however, still like the numbers to count when one can save either a lesser or greater number from equal or similar harm. What is therefore needed is a moral theory that allows the relative numbers to count sometimes but not always. Several philosophers have proposed theories of this kind.² The different approaches are motivated by a powerful idea: our decision whom to save should respect each person's separate

¹ While the opposition to full aggregation is a central feature of one prominent type of non-consequentialism, we should not confuse non-consequentialism simpliciter with the opposition to full aggregation. One can be opposed to full aggregation as part of one's theory of the good (e.g. Temkin, *Rethinking the Good*, ch. 3) while still being a consequentialist about rightness. Similarly, forms of non-consequentialism are not opposed to aggregation. For example, some non-consequentialists only depart from consequentialism by accepting either deontological constraints or agent-centered prerogatives (e.g. Samuel Scheffler, *The Rejection of Consequentialism* (Oxford: Oxford University Press, 1982). These forms of non-consequentialism, however, are only partial departures that tame a basically consequentialist outlook of morality with additional considerations. It is a half-hearted form of non-consequentialism (see Thomas Sinclair, "Are We Conditionally Obligated to be Effective Altruists?," *Philosophy & Public Affairs* 46 (2018): 36-59, at pp. 43-49). The opposition to full aggregation which is subject of this chapter is instead the core of a more thorough form of non-consequentialism.

² Kamm, *Mortality, Mortality*, 1:156-61; *Intricate Ethics*, pp. 31-77; Scanlon, *What We Owe to Each Other*, pp. 238-41; Lefkowitz, "On the Concept of a Morally Relevant Harm"; and Voorhoeve, "How Should We Aggregate Competing Claims?". In this chapter, I am concerned with limited aggregation as a view about what we ought to do and not as an axiological principle. For the axiological version of limited aggregation see Temkin, *Rethinking the Good*, ch. 3 and Dale Dorsey, "Headaches, Lives and Value," *Utilitas* 21 (2009): 36-58.

claim to our help; in particular it should respect those in need whose claims are the greatest. Such views have been called *limited aggregation*. In this chapter, I will set out my own view of limited aggregation and show how such a view can be both intuitively plausible and respect the demands of the separateness of persons.

The standard cases for limited aggregation are cases in which groups are homogenous; i.e. groups in which everyone in the group has a claim that is as strong as the claim of every other member of the group. However, many cases are not like this. Oftentimes we face decisions where the groups are heterogenous; i.e. not all groups members face the same plight. Current proposals of limited aggregation have been shown to have devastating flaws when they are extended to cases with such heterogeneous groups. Patrick Tomlin has shown that applying a leading proposal of limited aggregation, *Aggregate Relevant Claims*, to heterogenous group cases violates one of two uncontroversial principles. On one extension of *Aggregate Relevant Claims* it violates a principle he calls *Equal Consideration for Equal Claims* which requires us to give all claims of equal strength equal weight in determining whom to save. On its alternative extension *Aggregate Relevant Claims* violates what he calls the *Principle of Addition*, which requires that adding a claim to a group cannot make saving this group less choiceworthy.³ In this chapter, I show how these problems can be resolved by a new theory of limited aggregation that is well-grounded in the reasons we have to be skeptical of aggregation in the first place and that meets this challenge set by Tomlin and related recent challenges. I propose the following theory:

Hybrid Balance Relevant Claims. Relevant individual claims ought to be balanced against one another, starting with the strongest claim(s) overall.

³ Patrick Tomlin, "On Limited Aggregation," *Philosophy & Public Affairs* 45 (2017): 232-60. Tomlin's criticism has been extended by Joe Horton, "Always Aggregate," *Philosophy & Public Affairs* 46 (2018): 160-74. An earlier line of criticism objected that limited aggregation violates axioms of rational choice, namely transitivity and the independence of irrelevant alternatives. See Derek Parfit, "Justifiability to Each Person," *Ratio* 16 (2003): 368-90, at pp. 384-85; and John Halstead, "The Numbers Always Count". Since this criticism has been, in my view, adequately responded to I will not address it except where it serves as a useful comparison to my arguments. For the responses see Kamm, *Intricate Ethics*, 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* 10 (2017): 148-56, at pp. 152-53; and Tomlin, "On Limited Aggregation," p. 236fn11.

If there are unbalanced relevant claims, then these will be decisive in what we ought to do. If the claims are evenly matched, then we are permitted to save either group, or perhaps required to give equal chances. The relevance of claims is determined by two conditions:

- (1) The local relevance condition: A claim can only be balanced with another claim if the two claims are relevant to one another.
- (2) The global relevance condition: Every individual with a strong claim has a veto against the consideration of any type of claim that is irrelevant to her claim, if considering these claims would lead to her not being saved.

In Section II, I will establish that the idea of relevance is key to any plausible theory of limited aggregation. Then, in Section III, I will provide my defense of Hybrid Balance Relevant claims. I will explain more precisely what I mean by “balancing” claims against one another and I justify the hybrid character of having both a local and global relevance condition. Given the complexity of Hybrid Balance Relevant Claims, I offer some illustration in Section IV before discussing how my view escapes all recent challenge that have been raised against limited aggregation in Section V.

II. Relevance and Limited Aggregation

I have already mentioned that many philosophers are opposed to aggregation because it allows a large number of trivial claims to outweigh the significant claim of a single individual. A paradigm case for this phenomenon is *Life versus Headaches*. A fully aggregative view struggles to accommodate the intuition that we should not let the single person die. It seems that if we can aggregate the pain of the various headaches, there will be a number of people suffering from headaches that outweighs the life of the one.⁴

⁴ Of course, some nonetheless defend full aggregation. See Norcross, “Comparing Harms”; Hirose, *Moral Aggregation*, chs. 2-3; Horton, “Aggregation, Complaints, and Risk”; and Horton, “Always Aggregate”. Fully aggregative views differ with respect to what ought to be aggregated. Norcross believes we ought to aggregate harms, Hirose believes in formal

There are some views that allow aggregation in all cases but seek to avoid this conclusion. For example, a view could be fully aggregative but assign infinite disvalue to deaths. Such a view would then, however, struggle to accommodate the intuition that sometimes the relative numbers should count. The disvalue of two deaths would also be infinite and no number of a slightly lesser harm than death could outweigh single deaths.⁵ Another attempt to accommodate the intuition in *Life versus Headaches* is to argue that we should accept full aggregation while adding that value functions are bounded. In such bounded value functions, the aggregate value of any number of a given harm has an upper bound. As the number of headaches approaches infinity, the value of saving these people approaches a fixed value lower than the value of saving a single person from death.⁶ Such views imply, implausibly, that our reasons for saving persons from serious harm diminish with the number of affected people. At some large number n , we have virtually no reason whatsoever to save additional people from serious harm. Such a view would therefore imply that we should rather save n people from a severe disability alongside one person with a headache instead of saving $n+1$ persons from a severe disability.⁷

The better solution is, therefore, to adopt the idea that there are different kinds of claims. There is something that distinguishes headaches from deaths in a manner

aggregation which can integrate a variety of moral factors, Horton believes we should aggregate complaints. Arguably Liao also falls into this category, though some remarks about irrelevant utilities seem to indicate otherwise. See S. Matthew Liao, "Who Is Afraid of Numbers?," *Utilitas* 20 (2008): 447-61.

⁵ See also Otsuka, "Saving Lives, Moral Theory, and the Claims of Individuals," pp. 127-28.

⁶ See Seth Lazar and Chad Lee-Stronach, "Axiological Absolutism and Risk," *Noûs* 53 (2019): 97-113. For objections similar to the one I raise see Otsuka, "Saving Lives, Moral Theory, and the Claims of Individuals," p. 127fn31 and Alex Voorhoeve, "Balancing small against large burdens," *Behavioural Public Policy* 2 (2018): 125-42, at pp. 132-34.

⁷ A proponent of bounded value functions might argue that not all value functions for saving people from harm are bounded. Perhaps the value of saving persons from death and all conditions which can be traded off against death is not bounded. The problem for this reply is two-fold. First, it would need to explain why it is the case that only for some conditions the value of saving additional persons diminishes. This appears like an unjustified restriction made only to avoid counterexamples. Second, the above argument would still hold provided that the severe condition is such that it cannot be traded off against death. For example, assume no number of broken legs may outweigh a single death. Then the value of saving people from broken legs is bounded. This would imply that while no number of mild headaches can outweigh a single broken leg, saving $n+1$ persons from a broken leg can be outweighed by saving n persons from a broken leg and a single person from a mild headache.

that bars us from trading off lives against headache relief. The idea of relevance can help here.⁸ Claims to headache relief are not relevant to claims to be saved from death. The same idea can explain why when the harms are more similar it is that the numbers should count. Consider *Life versus Paraplegia*. If we save whichever group has the strongest individual claim, disregarding the numbers entirely, then we should save a single person from death regardless of the number of people that we could save from paraplegia.⁹ However, *Life versus Paraplegia* is different from *Life versus Headaches* insofar as the claims to be saved from paraplegia are, plausibly, relevant to the claims to be saved from death.¹⁰

While the idea of relevance can give us a principle that can explain our intuitions in cases like *Life versus Paraplegia* and *Life versus Headaches*, the idea is incomplete in two ways. First, it is unclear how the idea of relevance can plausibly be extended to more complicated cases including those involving heterogeneous groups. Second, even if we had a decision procedure for such cases the question remains how to theoretically justify this decision procedure. While intuitive fit is an important part

⁸ See Kamm, *Morality, Mortality*, 1:144-97 and Lefkowitz, "On the Concept of a Morally Relevant Harm".

⁹ Throughout the chapter, I am using the term "group" liberally and sometimes refer to single individuals as a group.

¹⁰ Again, some disagree and think that in both cases we ought to save the single person from death. Most of those who disagree embrace the further claims that there is no obligation to save the greater number even when the harms are identical. See G.E.M. Anscombe, "Who Is Wronged? Philippa Foot on Double Effect: One Point, in *Elizabeth Anscombe, Human Life, Action and Ethics*, ed. Mary Geach and Luke Gormally (Exeter: Imprint Academic, 2005), pp. 249-51; Taurek, "Should the Numbers Count?"; and Tyler Doggett, "Saving the Few," *Noûs* 47 (2013): 302-15. Munoz-Dardé ("The Distribution of Numbers and the Comprehensiveness of Reasons") can be interpreted to support this position as well. As with full aggregation, these opponents of aggregation form a broad tent. Munoz-Dardé believes that while there is no duty to save the greater number, aggregative reasons can be intelligible reasons to act upon. Taurek seems to deny this and advocates giving equal chances. Anscombe and Doggett only argue for the permissibility of saving the few which indicates that they are not opposed to saving the many. However, it is possible to accept a duty to save the greater number and treat both *Life versus Paraplegia* and *Life versus Headaches* alike. Scanlon outlines an argument before embracing a form of limited aggregation in *What We Owe to Each Other*, pp. 230-38. Otsuka canvasses a variety of arguments for the duty to save the greater number which do not imply that we can trade off paraplegia against lives in "Saving Lives, Moral Theory, and the Claims of Individuals," pp. 118-24. R. Jay Wallace advances such an argument for the duty to save the greater number without noting that his argument does not extend to cases in which lesser but relevant harms are at stake. See R. Jay Wallace, *The Moral Nexus* (Princeton: Princeton University Press, 2019), pp. 215-19.

of a good moral theory, we also need to give a justification for the theory. Otherwise our theory merely summarizes rather than justifies our immediate reactions about cases. This is the task I set out to do in the next section.

III. Justifying Hybrid Balance Relevant Claims

The starting point for my view is an idea that appears in the work of Thomas Nagel. Nagel writes about the reconciliation of two standpoints, the impartial and the partial standpoint. Impartial concern is, however, non-aggregative. Impartiality should not be confused with impersonality which is aggregative.¹¹ Impartiality is based on the recognition that everyone's life, including one's own, has objective importance and significance. Realizing this, we extend this significance to the lives of others. We imagine ourselves to be in their shoes and extend an impartial concern for them. This impartial concern is fragmented. It is divided between the different individuals. Unlike fully aggregative theories that interpret impartiality as impersonality, we are not eroding the distinction between different viewpoints. This fragmented concern takes seriously the separateness of persons.¹² The realization of the objective significance of everyone's life need not, however, lead us to abandon our own personal perspective in the world. Impartial concern goes along with legitimate partial concern for oneself.

The ideal that we are striving towards is unanimity. It is not unanimity in our rational self-interest, but rather unanimity among persons committed to finding common principles guiding our interactions. Our actions should be justifiable to each and every one who is affected. There are different models of unanimity. One model of unanimity is the kind of unanimity that is achieved behind a veil of ignorance.

¹¹ Nagel, *Mortal Questions*, ch. 8; Nagel, *Equality and Partiality*, chs. 2-8. Unlike me, Nagel sometimes speaks of impersonal concern as interchangeable with impartial concern. Impersonality might be one way to show impartiality, but it is not the only one. In Chapter 2 (*Separate Persons Behind the Veil*) I discussed the relation between impartiality and impersonality or monoperpersonality. I argued that respect for the separateness of persons dictates that we should not equate impartiality with impersonality. The contrast is also drawn by Rawls in *A Theory of Justice*, rev. edn., pp. 165-68.

¹² Nagel, *Mortal Questions*, pp. 126-27.

Such a veil of ignorance achieves unanimity only by assimilating persons and depriving them of their separateness. By contrast, the kind of unanimity we are searching for is unanimity that is achieved by convergence from different standpoints.¹³

The two different perspectives, partial and impartial, can explain which claims are relevant.¹⁴ From our first-personal perspective we have a justified stronger concern for our own life than for the lives of others. When we imagine ourselves to be in the position of others, such imagination includes their self-favoring concerns. While individuals are entitled not to make use of their self-favoring concern, we cannot assume that individuals have volunteered to waive their moral claim to be aided. In the absence of any such waiver, we need to assume that individuals aim to promote their own perspective to the maximum extent that is morally permissible. The maximum extent of such concern determines when claims are relevant. A claim is relevant to another claim if and only if this claim can be preferred to the other claim from someone's point of view. A claim might be weaker and still relevant if the claim can be preferred from someone's self-favoring, partial perspective.

This explanation fits best with a non-welfarist understanding of moral claims.¹⁵ The ultimate question is whether individuals are allowed to have a stronger concern to save themselves from a given harm rather than someone else being saved from a greater harm. Our judgments about the relative priority or urgency to come to the rescue of persons do not always track the judgments of the people benefited about what makes their life go best. A person may believe that her wish to play all of Beethoven's sonatas is more important than a decent diet. But this does not make it the case that our reasons to aid her in the endeavor of learning the sonatas are stronger than our reasons to aid her with nutrition.¹⁶ If this is so, we should not

¹³ Nagel, *Equality and Partiality*, pp. 33-40.

¹⁴ I borrow this argument from Voorhoeve, "How Should We Aggregate Competing Claims?"

¹⁵ Here I differ from the way Voorhoeve sets up his view. His *Aggregate Relevant Claims* gives an analysis of claims in terms of the contributions to well-being that helping the person would make. "How Should We Aggregate Competing Claims?," pp. 64-66.

¹⁶ The idea originates with T.M. Scanlon, "Preference and Urgency," *Journal of Philosophy* 72 (1975): 655-69. The example is from Nagel's further development of this idea in terms of person-neutral and person-relative reasons in *The View From Nowhere* (New York: Oxford University Press, 1986), pp. 166-75.

believe that we can explain the strength of individual claims to be saved in terms of the contribution that saving the person will make to her well-being.

The relevance test that invokes the partial perspective means that in a case where the claims of only one of the groups are relevant, unanimity naturally emerges. We can see this with our paradigmatic case *Life versus Headaches*. From the perspective of the person about to lose her life, she should be favored. Both her impartial and partial concern favor this. From the perspective of each of the persons about to suffer a headache, the rescue of the life should be favored too. Their partial concern does not extend to the saving of themselves from a minor headache rather than another person's life. It would be unreasonable for these people to insist on their claim to be rescued. We can give a powerful justification to them for not rescuing them. Not even you with your partial concern would have been allowed to rescue yourself. So, how can you complain to me who does not have partial concern for you, for not rescuing you?

More difficult are those cases in which claims of equal strength are at play, but the relative numbers differ. For example, we can save either A or B&C from equal harm. In this case every person has a claim that survives the test of partial concern. Everyone can legitimately stake their claim to rescue. This means that we have a conflict of different standpoints. How should we resolve this conflict of standpoints? I endorse a method that I call *Balance Relevant Claims*. Balance Relevant Claims resolves conflicts of standpoints by balancing individual claims against one another.¹⁷ Consider the following pair of cases. In the first case we can either save A or B, in the second case we can either save A or B&C. While in the first case the considerations for saving A and the considerations for saving B are equally strong, this is not the

¹⁷ Proponents of views similar to Balance Relevant Claims are Kamm and Scanlon. See Kamm, *Morality, Mortality*, 1:101, 114-19; *Intricate Ethics*, pp. 31-77; *Bioethical Prescriptions* (Oxford: Oxford University Press, 2013), pp. 367-71, 515-22; and Scanlon, *What We Owe to Each Other*, pp. 231-35, 240-41. Views similar to Balance Relevant Claims are also called "local relevance" see Victor Tadros, "Localized Restricted Aggregation," in *Oxford Studies in Political Philosophy. Volume 5*, ed. David Sobel, Peter Vallentyne, and Steven Wall (Oxford: Oxford University Press, 2019), pp. 171-204; and Aart van Gils and Patrick Tomlin, "Relevance Rides Again? Aggregation and Local Relevance," in *Oxford Studies in Political Philosophy. Volume 6*, ed. David Sobel, Peter Vallentyne, and Steven Wall (Oxford: Oxford University Press, 2020), pp. 221-55.

case in the second case. The fact that C is a third person means that C can break the tie and decide that we should save B&C. We can explain the tie-breaking idea in terms of balancing claims. When we see that we can save A from death, we notice a strong claim on our help. If A's claim were the only thing to consider, we would be required to save A. But B's presence and B's claim balances the moral claim that A can raise. B's claim is just as strong as A's. Neither claim can ultimately decide what we ought to do. Since we are required to save someone, this means that either option is permissible, or perhaps we are required to give equal chances to both. But when C enters the picture, C's claim is not balanced. C's claim can then have the power to ultimately decide what we ought to do: namely, save B&C.

The tie-breaking idea helps us to better understand how Balance Relevant Claims works, but it is not a good guide to justifying it. One unsuccessful justification for Balance Relevant Claims involves an appeal to the moral complaint that if we are not required to save B&C, then the additional presence of C does not make a moral difference. However, the additional presence of C can make a moral difference in other ways. For example, a weighted lottery that reflects the different numbers would ensure that the additional presence of C makes a moral difference by shifting the odds.¹⁸ For this reason, we should not rest our case for Balance Relevant Claims on the idea that tie-breaking is the only way to respect the fact of C's additional presence. Instead, Balance Relevant Claims is justified holistically. Its justification depends on the prior acceptance of limited aggregation and the idea of relevance for which I argued in the previous section. It is then justified as a plausible method of resolving conflicts of standpoints in a manner that explains the intuitions that limited aggregation wants to capture. My aim in the following discussion is to make clear

¹⁸ Both Kamm and Scanlon rest their argument for Balance Relevant Claims on this moral difference argument. See Kamm, *Morality, Morality*, 1:101, 114-19; *Intricate Ethics*, pp. 31-32; and Scanlon, *What We Owe to Each Other*, pp. 231-35. The objection that one can make a moral difference in other ways is due to Michael Otsuka ("Saving Lives, Moral Theory, and the Claims of Individuals," pp. 114-18). In their analysis of Kamm and Scanlon, David Wasserman and Alan Strudler helpfully distinguish between what they call the marginal difference argument and the balancing argument. They clarify that Kamm and Scanlon think that the marginal difference argument grounds the balancing argument. My point here is that we can retain the balancing argument by giving it fresh and better foundations. David Wasserman and Alan Strudler, "Can a Nonconsequentialist Count Lives?," *Philosophy & Public Affairs* 31 (2003): 71-94, at pp. 82-89.

that Balance Relevant Claims can be developed into an overall compelling theory of limited aggregation.

Before proceeding, one further clarification is needed. We should distinguish carefully between balancing and canceling.¹⁹ It is misleading and incorrect to say that the claims of A and B cancel each other out and can thus be ignored. We can see this in the case where only A's and B's fate is at stake. Even though the claims balance each other, we are required to save one of the two. The claims are not canceled and thus remain within our moral deliberation. Not saving either of the two is impermissible. The fact that the claims are balanced only means that they do not have the force to *ultimately* decide which course of action is required.

This case of balancing highlights and expresses a different model of unanimity. In the case of *Life versus Headaches* we achieved unanimity by individuals withdrawing their claims because their partial perspective did not allow them to stake their claim. In the present case all claims are considered and weighed. We engage in a genuine confrontation of standpoints. But we can still say the following. It would be unreasonable for A to insist that his claim ultimately decides that she should be saved in the case of A versus B&C. Her claim was considered and balanced with the claim of someone else. She cannot insist that her claim has greater force than balancing a single claim. She can only insist on a fair decision procedure that takes her moral claims and the severity of her plight into consideration. Balancing meets this demand. But she cannot insist on more. It would be unreasonable for her to insist on a particular outcome in which she is saved and to reject a principle requiring B&C to be saved.²⁰

¹⁹ In the first version of her argument Frances Kamm did speak about canceling ("Equal Treatment and Equal Chances," *Philosophy & Public Affairs* 14 (1985): 177-94). Later on, Kamm admits that the canceling metaphor is misleading (*Morality, Mortality*, 1:116-17). Rahul Kumar used a neutralizing metaphor in later work ("Contractualism on saving the many," *Analysis* 61 (2001): 165-70). Kumar's argument was criticized on similar grounds to the ones presented here by Michael Otsuka ("Saving Lives, Moral Theory, and the Claims of Individuals," pp. 118-19).

²⁰ This idea forms part of Scanlon's latest revision of his contractualism in response to the problems of aggregation. See Scanlon, "Contractualism and Justification". A similar idea is brought forward by Munoz-Dardé. See Munoz-Dardé, "The Distribution of Numbers and Comprehensiveness of Reasons," pp. 208-15.

If we accept the balancing of claims, the question arises whether we can extend the idea of balancing to cases that are different from tie-breaking cases. For example, can several claims to be saved from paraplegia balance a single claim to be saved from death?²¹ It seems plausible that sometimes a greater number of weaker claims can balance a lesser number of stronger claims. Having a relevant claim means that one's own partial perspective allows one to maintain one's claim and to insist that one's claim will be considered. All cases involving opposing relevant claims thereby become conflicts of standpoints in which multiple people can rightfully insist that their claims are considered by the decision-maker. While they can all reasonably insist that their claims are considered, no one can insist that their claim has to be decisive. Balancing is a general method that allows us to resolve cases of such conflicts of standpoints. If one claim is outweighed by a multitude of weaker claims, this person was outweighed by people who were entitled to voice their claims and have their claims considered. If one's claim has been considered and was outweighed, it is unreasonable for single individuals to insist that their claim has to have absolute priority over all weaker claims. If we accepted the absolute priority of stronger over weaker claims, there would be little point in distinguishing between relevant and irrelevant claims. However, the view that I advance here draws a distinction between cases in which claims cannot be staked because they are irrelevant, and cases in which claims are considered and outweighed.

The next extension of the idea of balancing concerns cases with heterogenous groups. How should we decide which claims are relevant in cases in which not all members of the group have claims of equal strength? And how should we decide in which order to balance claims against one another? In cases like *Life versus Headaches* or *Life versus Paraplegia*, the questions of relevance and order are more straightforward. Either the claims that compete with the claims to be saved from death are relevant to the claim to be saved from death, or they are not. Given that all opposing claims are of equal strength, there is no question of different orders in which to balance claims either. Things are more complicated in heterogenous group

²¹ Wasserman and Strudler urge that they should. Wasserman and Strudler, "Can a Nonconsequentialist Count Lives?," pp. 89-92.

cases. However, I believe that we can apply the same principle that explains the difference between *Life versus Headaches* and *Life versus Paraplegia* to the case of competing heterogeneous groups.

Not saving a person with a strong claim to be saved requires a special justification to this person. Given that the person with the strongest claim is most likely to have grounds for grievance or complaint against our failure to save her, our justification for acting as we do must be primarily addressed to those who have the strongest claim on our aid. This gives an answer to the question of the order of balancing. It explains why balancing starts with the strongest claim and then works its way down to less strong claims. The stronger the claim the greater is the urgency to give a justification to this person.²²

The same idea can also illuminate when claims are relevant. The justification we can give to this person has to respect the claim of the person to assistance. In *Life versus Paraplegia*, we can point out to the person dying that the claims of the many are all relevant to her plight. The other individuals are entitled to stake their claims on us. By contrast, it would be disrespectful to deny saving a person from death by pointing out that doing so would enable the agent to prevent some number of headaches. To consider the headaches as a reason not to save the person from death would trivialize her situation. The headaches are simply not relevant in comparison to the death.²³ To say to the person: "I am sorry, but we cannot save your life because we are busy preventing many minor headaches" would trivialize what she stands to lose. This is not the case when we say: "I am sorry, but we cannot save your life because we are busy preventing many people from becoming paraplegic". The "because" clause in the justification indicates that whether or not a justification is

²² A possible alternative way to recognize the priority of the strongest claims is to balance in the interest of the person with the strongest claim. For a discussion of this possibility see van Gils and Tomlin, "Relevance Rides Again?," pp. 250-52. Fortunately, we need not decide between the two principles. In my appendix "The Order of Balancing" I argue that balancing in the interest of the person with the strongest claim is equivalent to my Hybrid Balance Relevant Claims view except in cases where balancing in the interest of the person with the strongest claim violates Equal Consideration for Equal Claims.

²³ For this see Kamm, *Intricate Ethics*, pp. 297-98, 484-87; *Bioethical Prescriptions*, pp. 368-71; Lefkowitz, "On the Concept of a Morally Relevant Harm," pp. 421-23; and Voorhoeve, "Why One Should Only Count Claims with which One Can Sympathize," pp. 152-53.

respectful depends on for whose sake we fail to save the person from death. This suggests the following general principle:

Respectful Failure to Save Principle. Every person that we fail to save is entitled to a respectful justification for our failure to save. It is disrespectful and impermissible to fail to save a person with a strong claim for the sake of persons whose claims are irrelevant to this strong claim.

The *Respectful Failure to Save Principle* tells us that what should guide our thinking about whether or not claims are relevant is determined by considering whether we can give a respectful justification to the person whom we fail to save. Applied to the case of balancing the claims of heterogenous groups, this principle identifies two scenarios in which counting claims as relevant renders us unable to respectfully justify our refusal to save.

The first scenario concerns a local feature of the confrontation of two claims. Whether or not we can justify the balancing of claims will depend on the relation between the claim that is balanced and the balancing claims. These claims will need to be relevant to one another in order for the balancing to be respectful. Balancing a claim means depriving it of the force to ultimately decide what one ought to do. The claim still has force in deciding what ought to be done; in this sense it is not canceled. But it loses the force to ultimately “tip the balance”. This is quite serious and deserves justification. We cannot justify it to a person that the moral force of her claim is diminished because of claims that are irrelevant to her claim. It is part of the meaning of an irrelevant claim that if a claim is irrelevant to claim X, then it can neither diminish nor override claim X.

The second scenario is different. It highlights the existence of what Frances Kamm has called “irrelevant utilities”.²⁴ To consider a trivial harm like a headache or a sore throat as the ultimate reason to save a group would be trivializing the fate of a person whose life will not be saved and thus be disrespectful to her. The headache or sore throat is not relevant for this decision. This type of disrespect that flows from the *Respectful Failure of Save Principle* is rather a global feature of the entire trade-off. It is

²⁴ Kamm, *Morality, Mortality*, 1:144-64.

disrespectful not because of the confrontation of two claims but rather because counting the trivial harm is disrespectful to the person whose life is at stake. The strongest claim in the opposing group fixes globally that trivial harms do not count.

What this means is that in each case of balancing, the claim must be relevant to the claim that is being balanced. As long as a person's claim is relevant to the fate of the other person, this person can rightfully insist that her claim ought to be taken seriously. But if her claim is irrelevant, then the person would have to withdraw her claim and cannot insist that her claim needs to be balanced against the competing claim. This is the first part of my view. However, there is a second condition. What about claims that are locally relevant and would be ultimately decisive? Suppose, for the sake of argument, that claims to be saved from minor headaches would ultimately determine that we ought not save a person from death. From her personal perspective the person facing the minor headache would not be allowed to save herself if someone else would otherwise die. She would not be allowed to uphold her claim if her claim was only competing with the claim of the person about to die. So why should this person then be allowed to uphold her claim if doing so results in failing to save the dying person in the more complex case where more claims are at stake? In both cases, we would fail to save someone from death for the sake of another person whose claims are not relevant to the person whom we fail to save. This cannot amount to a satisfactory and respectful justification to the dying person. Therefore, the claims against headaches should not be counted as relevant.

This argument leads to the following two conditions for determining the relevance of claims:

- (1) The local relevance condition: A claim can only be balanced with another claim if the two claims are relevant to one another.
- (2) The global relevance condition: Every individual with a strong claim has a veto against the consideration of any type of claim that is irrelevant to her claim, if such consideration will lead to her not being saved.

The first scenario illustrates the need for a local relevance condition, while the second scenario illustrates the need for a global relevance condition. The two relevance conditions also illustrate and underline how balancing is distinct from

ordinary aggregation. Balance Relevant Claims contrasts here with *Aggregate Relevant Claims*.²⁵ *Aggregate Relevant Claims* resolves conflicts of standpoints by appealing to aggregation. It thereby allows for aggregation, but only in a subset of cases. Only in those cases in which there is a conflict of standpoints, i.e. only in those cases in which the opposing claims are relevant to one another, can we permissibly aggregate. For *Aggregate Relevant Claims* it is clear from the outset which claims are relevant and hence can permissibly be aggregated. It is a crucial aspect of this view that the reason why we ought to save one group rather than the other is that the sum-total of claims is greater. This form of limited aggregation has a greater affinity with full aggregation.

By contrast, for Balance Relevant Claims we need to reason sequentially and match individual claims against one another in order to determine which claims are relevant. The process of balancing makes it possible to identify which opposing claims deprived a person's claims to be rescued. Balancing also allows us naturally to mention for whose sake we fail to save someone. It thereby sets clearly and intuitively apart cases where the ultimate reason for failing to save a person is a claim of the same magnitude, a relevant magnitude or an irrelevant magnitude. This means that we need to depart from a model that determines whether claims are relevant or not at the outset and then proceeds to aggregate those claims that are relevant. *Aggregate Relevant Claims* could not distinguish for whose sake we fail to save a person if different kinds of claims are allowed to be aggregated.

This difference in determining the relevance of claims further indicates that the method of counting claims employed by Balance Relevant Claims is subtly different from the method employed by *Aggregate Relevant Claims*. The following analogy can help bring this out. There are two ways in mathematics to determine whether one set is larger than the other. One way counts the members of the set and then compares the number of elements in the set. Here the cardinal number, or sum-total, matters. Another way of comparing the size of sets does not require any numeracy skills or even knowledge of numbers. We can see whether there is a

²⁵ A proponent of *Aggregate Relevant Claims* is Voorhoeve, "How Should We Aggregate Competing Claims?".

bijection, a one-to-one correspondence mapping between all elements of the two sets. In this method we only need to map individual members against one another. Balancing is like this second approach in that it maps claims against one another rather than counting a sum-total. The way balancing counts claims employs a similar form of reasoning as the anonymous Pareto principle does. Like Paretian reasoning, balancing places the claims of different groups in one-to-one correspondence relations and observes whether one group ends up with stronger claims than the other. B&C is anonymously Pareto superior to A because while B's claim can be matched by A's claim, C's claim cannot be so matched. The logic of the anonymous Pareto principle departs from simple aggregation.²⁶ Balancing extends a similar reasoning beyond the case of tie-breakers.

The difference between the way Balance Relevant Claims counts claims and the way Aggregate Relevant Claims counts claims highlights that it would be a mistake to lay too much emphasis on the question whether or not Balance Relevant Claims is aggregative in some sense. For example, in "Saving Lives, Moral Theory, and the Claims of Individuals" Otsuka discusses and rejects ideas about balancing *inter alia* because they do not abide by the *individualist restriction*, i.e. the constraint that the justifiability of moral principles depends only on that principle's implications for single individuals.²⁷ I advocate here that we should not think that all ways of violating the individualist restriction are equally bad. While both Aggregate Relevant Claims and Balance Relevant Claims seek to limit the role of aggregation, Aggregate Relevant Claims does so by appealing to a form of *restricted aggregation* while Balance Relevant Claims seeks to introduce a form of *aggregation light*. To better respect the separateness of persons, Balance Relevant Claims proposes a different way of reasoning about cases in which the standpoints of different individuals conflict.

Before proceeding, I address one concern about the hybrid character of my view. The worry is that it is gerrymandered and can explain our intuitions about cases only because it relies on divergent conditions that cannot be coherently defended. In

²⁶ See also Iwao Hirose, "Saving the greater number without combining claims," *Analysis* 61 (2001): 341-43.

²⁷ For the individualist restriction see Parfit, *On What Matters*, 1:193.

response, I have argued that both conditions originate from a single overarching principle. This principle illuminates the idea of respect that proponents of limited aggregation have relied upon. The fact that there are two conditions is only the reflection of the fact that disrespect can manifest itself in various ways. A related concern is that the combination of the two conditions does not fit well the model of balancing. The concern is that the conditions license a form of double counting. Individuals with a claim that is balanced can nonetheless exercise their veto against the consideration of other claims. This claim would have balanced some claims *and* vetoed other claims. In response, we can see more clearly that this is not double counting by reminding ourselves of the distinction between balancing and canceling. Balancing means that a claim is only deprived of some of its moral force. In a confrontation between two equally strong claimants, each of the claims will be balanced yet nonetheless we are required to save one of them. These balanced claims retain some moral force to ensure that we cannot escape our duty to save because of the fact that both are equally deserving of our aid. The reason why a balanced claim retains this power is that failing to save anyone in the case of equally balanced claims would disrespect their moral standing as persons who deserve to be saved. Similarly, without the veto condition, a decision-maker would be licensed to fail to save the person in a disrespectful manner. A person's claim can never be deprived of the force to insist on respectful treatment.

IV. Illustrations of Hybrid Balance Relevant Claims

The view that I defend is complex. I will help to illustrate it with three examples. We can imagine the examples to be decisions about the allocation of scarce medical resources that can be used to save people from permanent medical conditions and restore them to full health. The medical conditions are specified by broad categories of severe impairments, moderate impairments, and mild impairments. While the mild impairments are relevant to the moderate impairments and the moderate impairments are relevant to the severe impairments, mild impairments are not relevant to severe impairments. The use of general categories

leaves open room for disagreement about what conditions are relevant to one another. It also leaves open the possibility that two individuals have the same claim to aid even though one is facing a slightly worse hardship.²⁸ I take no stance on these issues here and will rather assume that the claims of individuals fall within these three categories. Hybrid Balance Relevant Claims is compatible with a variety of views concerning the relative moral importance of alleviating different conditions. For the sake of my illustrations we can assume that ten claims against a moderate impairment are equal in strength to one claim against a severe impairment. Likewise, ten claims against a mild impairment are equal in strength to one claim against a moderate impairment. In my examples I illustrate claims that are balanced on either side by bracketing them.

Table 4.1: Case One

Group A	Group B
(1 Severe)	
1 Moderate	(10 Moderate)
	11 Mild

In Case One the claim to be saved from a severe impairment is balanced by the ten claims against moderate impairment. If it was not for my global relevance condition, the claim against moderate impairment in favor of Group A could be outweighed by the claims against mild impairment. However, the global relevance condition blocks this. It would be disrespectful to the person with the severe impairment not to save her for the sake of people afflicted with a mild impairment. Considering the mild impairments vitiates a respectful justification that we can give to the person losing out.

Table 4.2: Case Two

Group A	Group B
(1 Severe)	
(10 Mild) + 1 Mild	(10 Moderate) + (1 Moderate)

²⁸ See Kamm, *Bioethical Prescriptions*, pp. 408-11.

Case Two illustrates when claims that are irrelevant to the strongest overall claim nonetheless remain relevant. Here the claims against severe impairment are balanced by ten claims against a moderate impairment. The moderate impairment would then indicate that we should save Group B. But B's claim can be balanced in favor of A by taking into consideration the claims of the mildly impaired. Ignoring the claims of the mildly impaired here would not be disrespectful towards the one with the severe impairment. On the contrary, it is in her interest that these claims are to be counted.

Table 4.3: Case Three

Group A	Group B
(1 Severe)	(10 Moderate) + (1 Moderate)
(10 Mild) + (1 Mild)	(1 Mild) + 1 Mild

But what about Case Three then? Here the last unbalanced claim is one of a mild impairment and this claim advocates not saving the group with the most serious claim of severe impairment. Can we justify this to the person with the severe impairment? I believe we can. Had we not counted the persons with mild impairments in this case, we should still save Group B. The claims in Group B to be saved from a moderate impairment would outweigh the claim of the single person with a severe impairment in A. The person with the severe impairment cannot complain that we do not save her because of the claims to be freed from the mild impairment. Had we not counted these we would still not be permitted to save her. If only claims against severe and moderate impairment count, we ought to save B. Accordingly, my global relevance condition does not block considering the claims of mild impairment in this case. There is no point for the person with the severe impairment to exercise her veto, since she would not be saved even if we fail to consider the claims of mild impairment. Consequently, we can justify our failure to save to the person with the severe impairment. Counting the persons with mild impairments does not vitiate a justification that we can give to this person.

Cases One and Two illustrate that the hybrid view has a certain asymmetry. We consider claims of mild impairment when they favor the person with the

strongest claim, but not when they oppose the strongest claim. This asymmetry may seem suspect. The asymmetry does not, however, violate the principle that Tomlin has called *Equal Consideration for Equal Claims*. According to *Equal Consideration for Equal Claims*, we should give equal weight to claims of equal strength. My view does this. Whenever some claims of a certain relevance class become relevant, all claims of that class become relevant. If, in Case Two, there were other claims against mild impairment on the side of Group B, these would equally have to be counted. For every moral choice, it is either the case that the claims of a given class are relevant or irrelevant.

The asymmetry is more modest than a violation of *Equal Consideration for Equal Claims* in that the relevance of less strong claims depends, inter alia, on which group has the strongest claim in its favor. This modest asymmetry can be defended. My previous argument that illustrates the different justifications we can give to the person with the strongest claim does just this. Counting mild impairments in Case One would make the failure to save the person with the severe impairment disrespectful. Counting the mild impairment in Case Two, however, would not be disrespectful to the persons in Group B whom we would fail to save. Further, for the reasons I outlined above, counting mild impairments in Case Three does not vitiate the justification we can give to the person with a severe impairment. If this explanation succeeds, then the asymmetry is justified.

One other feature may seem to violate *Equal Consideration for Equal Claims*. Consider the following two cases.

Table 4.4: Variation on Case Two

Group A	Group B
<i>Case Two</i>	
(1 Severe)	(10 Moderate) + (1 Moderate)
(10 Mild) + 1 Mild	
<i>Case Two*</i>	
(1 Severe) + (1 Severe)	(1 Severe)
11 Mild	(10 Moderate) + 1 Moderate

In Case Two, we ought to save Group A. The claims against mild impairment are locally relevant to the claims of moderate impairment in Group B. In Case Two*, we ought to save Group B. The claims against mild impairment are irrelevant according to my global relevance condition because of the presence of a claim against severe impairment in Group B. Even though we added equal claims to both sides, our verdict of permissibility changes. This feature does not violate *Equal Consideration for Equal Claims*. Both claims against severe impairment are equally considered. Indeed, it is because they are considered equally that claims against mild impairment become irrelevant on either side.

While it might seem counterintuitive that the addition of equally strong claims should change the permissibility of our decision, there is a rationale for this. With the addition of the claim against severe impairment in Group B something important changes. We can no longer justify considering the claims against mild impairment as ultimately decisive. In Case Two this was plausible since doing so would not result in a person not being saved whose claims are in a different class of relevance. But this changes in Case Two*. This change should lead us to consider the case differently and accept the verdict that Group B ought to be saved. The addition of the same claim on both sides can therefore change what one ought to do.

V. Hybrid Balance Relevant Claims and Objections to Limited Aggregation

So far, I have defended Hybrid Balance Relevant Claims as a theory of limited aggregation that can be applied both to homogenous group cases and to heterogenous group cases. It is also well-grounded in the commitment to the separateness of persons. In the remainder of the chapter I will show how the hybrid character of my view escapes challenges that other views of limited aggregation face. I will start with challenges to views that rely on a global relevance condition before moving on to challenges against purely local relevance. In a third step, I discuss and reject one challenge which all views of limited aggregation face.

A. Problems with Global Relevance

Consider first Tomlin’s challenge to limited aggregation. Tomlin has presented this as a problem against Voorhoeve’s version of Aggregate Relevant Claims. Tomlin has pointed out that the idea of relevance in Aggregate Relevant Claims is ambiguous between two interpretations. Are claims relevant if they are relevant to the strongest claim with which they compete? Or are claims relevant if they are relevant to the strongest claim overall? Tomlin calls this the Anchoring Problem and goes on to argue that in either case Aggregate Relevant Claims has deeply implausible implications when applied to heterogenous group cases. The problem with Voorhoeve’s Aggregate Relevant Claims is that it only contains a global relevance condition. The two alternatives that Tomlin outlines are both forms of global relevance that tell us that claims either are or are not relevant in a given choice situation. Since my view is a hybrid view that incorporates both a global and a local relevance condition, it avoids the Anchoring Problem.

The first possibility is that claims are relevant when they are relevant to the strongest claim with which they compete (*Anchor by Competition*).²⁹ Tomlin provides the following counterexample.

Table 4.5: Anchor by Competition

Group A	Group B
<i>Anchor by Competition Case One</i>	
1 Severe	10 Moderate
<hr style="border: 0.5px solid black;"/>	
<i>Anchor by Competition Case Two</i>	
1 Severe	10 Moderate
1 Mild	10 Mild
<hr style="border: 0.5px solid black;"/>	

In Case One the two groups are evenly matched. In Case Two, the claim of the member of Group A against a mild impairment is relevant because it is relevant

²⁹ Tomlin, “On Limited Aggregation,” pp. 240-44.

to the claims against moderate impairment in Group B. But the claims against a mild impairment in Group B are not relevant since they are not relevant to the claim against severe impairment in Group A. No matter how large the n , we will always favor Group A. This is deeply implausible. Not only is this implausible, it would also be very difficult to theoretically accept this conclusion. Anchor by Competition illustrates *Equal Consideration for Equal Claims*.

By contrast, my view resolves Case Two differently. In Case Two the global relevance condition treats the mild impairments as irrelevant utilities. This means that nothing changes for this case. We are still in a tie. The global relevance condition that I adopt respects *Equal Consideration for Equal Claims*. If a person with a strong claim can veto other claims, the person vetoes all claims of the same type. Thus, claims are always either relevant or irrelevant in a given situation.

Consider now Tomlin's second alternative (*Anchor by Strength*).³⁰ This alternative holds that relevance is determined by reference to the strongest overall claim, regardless of which side this claim favors. The following example illustrates this.

Table 4.6: Anchor by Strength

Group A	Group B
<i>Anchor by Strength Case One</i>	
111 Mild	11 Moderate
<hr style="border: 0.5px solid black;"/>	
<i>Anchor by Strength Case Two</i>	
1 Severe	11 Moderate
111 Mild	
<hr style="border: 0.5px solid black;"/>	

In Case One, the claims against mild impairment are relevant since the strongest overall claim is against moderate impairment. The claims against mild impairment can therefore outweigh the claims against moderate impairment. In Case Two, however, the claims against mild impairment are no longer relevant to the

³⁰ Tomlin, "On Limited Aggregation," pp. 244-47.

strongest overall claim. This makes it the case that the claims against moderate impairment can now outweigh the claim against severe impairment. Group B ought to be saved even though Group A has now one additional strong claim in its favor. This violates the *Principle of Addition* since Group A is less choiceworthy even though there is an additional claim in Group A present.

Anchor by Strength has the problem that only the strongest claim determines which claims are relevant. My hybrid view on the other hand allows that claims that are not relevant to the strongest claim can be relevant in balancing claims *provided that it is not to the disadvantage of any person with the strongest claim*. This means that the hybrid view, unlike Anchor by Strength, fulfils Tomlin's *Principle of Addition*. The violation of the *Principle of Addition* occurs when adding a claim can render less important claims irrelevant. But my view admits that such claims can still remain locally relevant and we are allowed to consider them when doing so does not weaken the case for the strongest claim. In the counterexample to Anchor by Strength it is even in the interest of the person with the strongest claim that the weaker claims are counted. The strongest claim cannot complain that considering these claims would be disrespectful to her. These claims are her "allies" and according to my view a strong claim cannot lose its "allies" by rendering them globally irrelevant. This means that the hybrid view, unlike Anchor by Strength, fulfils Tomlin's *Principle of Addition*.

Since my view implies neither of the two problematic case judgments and also does not violate either of the two principles that Tomlin proposes, it avoids the Anchoring Problem. As I have just shown, part of the reason for this is the acceptance of some form of local relevance condition. Tomlin himself, together with Aart van Gils, has proposed a view that embodies a form of local relevance as a possible response to the Anchoring Problem.³¹

³¹ Van Gils and Tomlin, "Relevance Rides Again?". Already in "On Limited Aggregation" Tomlin tentatively suggests a view of local relevance (pp. 259-60). He cites Garrett Cullity and Victor Tadros as inspirations for this kind of view. See also Tadros, "Localized Restricted Aggregation". I discuss the difference between Tomlin's (and van Gils's) version of balancing and mine in the appendix *Hybrid Balance Relevant Claims versus Sequential Matching*.

B. *Problems with Local Relevance*

Unlike Tomlin's (and van Gils's) own solution to the Anchoring Problem, I believe that while we should incorporate local relevance, we should not give up on global relevance altogether. To see why the hybrid character of my view is important consider Kamm's Sore Throat Case. Kamm's Sore Throat Case is an illustration of the problem of irrelevant utilities. In Kamm's case we have the choice between saving one life and saving another life alongside saving a person from a sore throat. But this additional sore throat should not tip the balance and render it mandatory to save the second person.³² Not all versions of limited aggregation are able to accommodate this case. In particular, this case presents a challenge to versions of Balance Relevant Claims that lack a global relevance condition. Such views cannot tell us why it is impermissible to count the claim of the sore throat.

My view that combines a local and global relevance condition can do so. In the Sore Throat Case the claim to be saved from a sore throat is globally irrelevant. It should not feature in our deliberation. The two claims to be saved from death would be evenly balanced against one another. Their force in ultimately deciding what we ought to do is thereby deprived. We are left with a tie. Either we have to give equal chances, or it is permissible to save either group.

The global relevance condition also allows my view to avoid a similar problem.³³ Imagine there is one claim against death in Group A that is balanced by enough claims in Group B so that a single claim against severe impairment remains. This claim against severe impairment is then balanced by enough claims in Group A so that a single claim against moderate impairment remains. This claim in turn is then balanced, and so on. In the end a single claim against a trivial inconvenience, like a sore throat, remains. If we did not accept a global relevance condition, we would have to accept that a sore throat could become decisive in this scenario. But it seems implausible that it really should. By accepting a global relevance condition, my view

³² Kamm, *Morality, Mortality*, 1:101-2, 146-47; *Intricate Ethics*, p. 34; and *Bioethical Prescriptions*, pp. 368-69.

³³ Van Gils and Tomlin, "Relevance Rides Again?," pp. 242-44.

can appealingly explain why the sore throat should not be decisive. It would be disrespectful not to save a person from death (or severe impairment for that matter) because of the existence of a sore throat. Under my view, instead of letting balancing proceed until trivial inconvenience such cases would be decided by claims that are still relevant to the strongest claim that we fail to satisfy.

Kamm's Sore Throat case is an objection to a view that adopts only a local relevance condition. While my view incorporates and explains the intuition in the Sore Throat case, accepting that the sore throat makes a difference may not be a decisive objection for a proponent of purely local relevance. Some philosophers writing on aggregation have expressed doubt whether we should retain the intuition that it is permissible to save either of the two persons in Kamm's case.³⁴ Versions of Balance Relevant Claims that contain only a local relevance condition are, however, subject to a different objection that my view avoids. Joe Horton provides this excellent criticism against views of purely local relevance. He devises the following case with two stages.³⁵

³⁴ See e.g. Campbell Brown, "Is close enough good enough?," *Economics & Philosophy* 36 (2020): 29-59, at pp. 41-42; van Gils and Tomlin, "Relevance Rides Again?," pp. 231-42; and Korbinian Rüger, "Aggregation with Constraints," *Utilitas* (forthcoming).

³⁵ For the case see Horton, "Always Aggregate," pp. 168-71. I changed the precise example so that it fits the stipulations about relevance that I have used throughout the paper. One further comment: Horton argues that the problem limited aggregation faces is a problem of path dependence. The order in which claims are balanced against one another matters (Horton, "Always Aggregate," pp. 167-68). The word "path dependence" is misleading, however. It indicates that the stages in Horton's case are sequential. But the temporal element introduces additional difficulties in the case. For if we knew in advance that the people in the later stage will be in this position, then we can simply skip the first choice. And if we did not know and only later come to know of the people that are "added", then this raises problems about the evidence-relativity of moral theories and about what we are required to do once we already made a commitment to helping some. To avoid these unnecessary complications, we should rather understand the different stages counterfactually. The objection then is that if there were additional people, then something implausible would follow.

Table 4.7: Horton against Local Relevance

Group A	Group B
<i>Local Relevance Stage One</i>	
1 Severe	
	1,000 Mild
<i>Local Relevance Stage Two</i>	
1 Severe	
11 Moderate	11 Moderate
	1,000 Mild

At Stage One, we should save the person from the severe impairment. What about Stage Two? A form of Balance Relevant Claims with only a local relevance condition would resolve it as follows: The claim against severe impairment in Group A can be balanced by claims against moderate impairment in Group B. The remaining claim against moderate impairment in Group B can be balanced against one claim against moderate impairment in Group A. The remaining ten claims against moderate impairment can also be balanced and even outweighed by claims against mild impairment in Group B. Hence, we ought to save Group B.

However, in this case intuitively we should still save Group A in Stage Two. The equal addition of claims should not make a difference in this case. While I argued that sometimes we can justify that the addition of equally strong claims on both sides makes a difference, it is hard to see why it should make a difference in this case. My argument was based on the idea that my adding a claim that is stronger than any other claim in the group, we change what justifications are available to this group for not saving them. This consideration is clearly not at stake in the present case. It would be desirable, therefore, if we could retain the judgment that in Stage Two we should save Group A as well. My view can do so in an intuitive fashion. The reason why we ought to save Group A in Stage Two as well is that the person with the severe impairment can complain and veto our failing to rescue her for the sake of people with mild impairments.

Horton explains that the failure of a view that only contains a local relevance condition is that it can allow for irrelevant claims to be “activated” in a manner in

which they can help outweighing the strongest claim. If there are any claims against moderate impairment in Group A, then, under the view Horton criticizes, it is always possible to balance the claims against moderate impairment in Group A against the claims against mild impairment in Group B.³⁶ Even if we added a greater number of claims against moderate impairment to Group A than to Group B, this could still mean that we should ultimately save Group B in Stage Two. The idea is that introducing claims against moderate impairment makes mild impairments relevant, so that they can contribute to our reasons for saving Group B. The claims against mild impairment could here outweigh the remaining claims against moderate impairment. But this is what my global relevance condition blocks. The person with a severe impairment has a veto against claims of mild impairment being considered. Mild impairments are treated as irrelevant utilities. My hybrid view therefore avoids the problem that Horton raises for views of pure local relevance.

C. *Principle of Agglomeration*

Horton, in his article, also raises another objection against limited aggregation. He describes the possibility that there are two separate moral decisions with one group we are required to save. But once we join the two decisions, it is possible that the group composed of those we ought to have saved no longer ought to be saved.³⁷ Horton might appeal here to something like the Principle of Agglomeration according to which combining groups that ought to be saved cannot render them a group that ought not to be saved.

The following case illustrates the problem.³⁸

³⁶ Horton, "Always Aggregate," pp. 171-73.

³⁷ Horton, "Always Aggregate," p. 173.

³⁸ Another, weaker, illustration is my previous argument that adding equal claims on both sides can make it the case that a different group ought to be saved. In this previous case merging a choice where A ought to be chosen with a choice where we ought to be indifferent can make it the case that B ought to be chosen. Here, it is the case that merging two decisions in which parts of A would be chosen makes it the case that B will be chosen.

Table 4.8: Principle of Agglomeration

Group A	Group B
<i>Combined Case</i>	
(1 Severe)	(10 Moderate) + (1 Moderate)
(10 Mild) + (980 Mild)	(980 Mild) + 20 Mild
<i>Sub-Case One</i>	
1 Severe	1,000 Mild
<i>Sub-Case Two</i>	
(110 Mild) + 880 Mild	(11 Moderate)

In both Sub-Cases we ought to save the sub-set of A, but in the combined case we ought to save the combination of the sub-sets of B. Notice, however, the following. We could also divide the case in the following sub-sets.

Table 4.9: Principle of Agglomeration (Alternative)

Group A	Group B
<i>Sub-Case Three</i>	
(1 Severe)	(10 Moderate) + 1 Moderate
<i>Sub-Case Four</i>	
(990 Mild)	(990 Mild) + 10 Mild

In these Sub-Cases it is obvious that any limited aggregation view would have to select B in both cases. What does it tell us? First, any view on limited aggregation needs to decide the four sub-cases the way I suggested. They follow straightforwardly from the stipulations I have made about relevance. If we accept the Principle of Agglomeration, we ought to save Group A because Sub-Case One and Two tell us to save sub-groups of A. But if we accept the Principle of Agglomeration,

we also ought to save Group B because Sub-Case Three and Four tell us to save sub-groups of B. Yet we cannot save both A and B. In other words, *any* view of limited aggregation has to violate the Principle of Agglomeration.³⁹

Agglomeration is intuitively appealing. What difference can *mere composition* really make? Should it really matter in which groups we are finding people assembled? Agglomeration shares one feature with the independence of irrelevant alternatives. Of course, if alternatives really are irrelevant then they should not count. The problem is that sometimes alternatives can be relevant even if they do not appear to be at first sight. Whether or not one option is on the table influences, for example, the kinds of justifications we can give to those we do not save. Similarly, composition can be relevant if it is not *mere* composition but something else that changes.

A successful agglomeration argument should be one where mere scaling up results in a reversal of permissibility. In Chapter 3 (*Contractualism, Complaints, and Risk*) I have argued that ex post contractualism does this. Ex post contractualism would be willing to impose objective and probabilistically independent risks on two distinct groups of people. But the very same risk imposition which affects each individual in the very same manner would be impermissible if imposed on the group as a whole. Here nothing of significance changes when we change the composition of the group.

This agglomeration argument differs crucially from Horton's. It is symmetrical in the sense that everyone is equally affected in the sub-groups. Horton's agglomeration cases do not have this feature. They crucially rely on the fact that the groups that are combined are not symmetrical. This makes it plausible that it is not *mere* composition that changes. Joining the groups together changes the moral situation by changing the relations of individuals towards one another. For example, the existence of persons with moderate impairments enables us to give the person with a severe impairment a justification we were previously unable to give. These special relations between different persons only arise in the particular configuration

³⁹ In his article, Horton claims that the agglomeration objection is one part of a dilemma for limited aggregation. What my argument here shows is that Horton is incorrect about this point. The concern about agglomeration is not a concern that any form of limited aggregation can avoid by embracing a different "horn".

of one case. They do not necessarily hold between sub-sets of these persons. Since limited aggregation is predicated on the idea that the moral relations of individuals to one another are important, we should not be surprised that changes in the relations of individuals will change permissibility verdicts. We can give a satisfactory explanation why agglomeration is not a mere change in composition.⁴⁰

VI. Conclusion

Limited aggregation is an attractive intermediate position between fully aggregative views and views that avoid all aggregation. But not all forms of aggregation are the same. I distinguished between two forms of limited aggregation, one with greater affinity to the outright rejection of all aggregation and another with greater affinity towards aggregative thinking. The motivation that leads us to reject full aggregation, the separateness of persons, is better developed by Balance Relevant Claims which show greater affinity to the outright denial of aggregation.

Hybrid Balance Relevant Claims can also refine our ideas about which claims should be treated as relevant and which ones should not. It imposes two requirements. One is local insofar as each individual instance of balancing has to occur between relevant claims. Another one is global insofar as it ensures that we can also give a respect-based justification to those whom we fail to save. We will never not save someone for the sake of people with claims that are irrelevant from the standpoint of those who we fail to save. This hybrid view is therefore not a cheap compromise between two theories that we ought to reject. Rather, the two components nicely flow from the common motivation that our decisions about whom to save should respect also those individuals whose claims we cannot fulfil.

If we adopt Hybrid Balance Relevant Claims we can respond to the recent criticism of limited aggregation. We can develop a view that avoids Tomlin's

⁴⁰ In his article Horton claims that his dilemma shows why intransitivity is not as innocuous as proponents of limited aggregation have claimed (Horton, "Always Aggregate," pp. 170-71). What my response here shows is that Horton's agglomeration problem is not any more troubling to limited aggregationists than previous challenges were. My reply here invokes the importance of moral relations and relational properties, the very same considerations that are invoked to argue why transitivity and the independence of irrelevant alternatives fail.

Anchoring Problem, as well as the problems that Horton has raised for views with purely local relevance. We only need to pay close attention to the guiding principle of respecting those whom we, sadly, cannot save.

Balancing Three (or More) Groups

In the chapter I have only considered cases with two distinct groups. My view should ideally not be limited to such cases but also possibly be applied to choices between three or more groups. This is not trivial. Balancing is by its nature confined to one-on-one comparisons. This might seem to limit or make impossible its applications to decisions with more than two groups.⁴¹

The easiest way to maintain this one-on-one confrontation is to engage in pairwise comparisons between the groups. We compare Group A to Group B, Group B to Group C, Group C to Group A. If one group wins both pairwise comparisons, then we ought to save this group. Pairwise comparisons cannot help us in all cases, however. Assume that the A-claims are against severe impairment, B-claims against moderate impairment, and C-claims against mild impairment. Without knowing the numbers, we can say that A ought to be saved rather than C. But we cannot say anything about the choices between A and B, and B and C. It is possible then that B ought to be saved rather than A, and C ought to be saved rather than B. If this is the case, we have a cycle and pairwise comparisons cannot tell us whom to save.

Balance Relevant Claims can, however, take a similar solution to this problem as Aggregate Relevant Claims.⁴² The guiding principle of Balance Relevant Claims is that we should be able to respectfully justify not saving the person with the strongest claim. In our three-option example, we know that we cannot justify saving C to the persons with a severe impairment. Ignoring their plight in favor of mild impairments would not give due respect to the severity of their condition. This means that C is not an option for us. Saving group A or B are both in principle possible. Neither would render the decision disrespectful to the person most likely to have a grievance. We can then compare A and B in isolation. Depending on the numbers we would save either A or B.

⁴¹ Wasserman and Strudler, "Can a Nonconsequentialist Count Lives?," p. 93.

⁴² For Voorhoeve's solution see Voorhoeve, "How Should We Aggregate Competing Claims?," pp. 76-78. Voorhoeve builds on the work of Frances Kamm. See Kamm, *Intricate Ethics*, pp. 297-98.

This method can also be applied to cases where the groups are heterogenous. Consider an example.

Table 4.10: Balancing Three Groups

Group A	Group B	Group C
1 Severe	11 Moderate	1 Moderate 101 Mild

In pairwise comparisons Group B is favored over A; Group C is favored over B; and Group A is favored over C. To break the cycle, we have to look at potential respectful justifications. Could saving Group A be justified to Group B? Yes, the person in Group A faces a more significant hardship than those in Group B. This can be a suitable justification. Could saving Group B be justified to Group C? Again yes, there are more people suffering from the moderate impairment in Group B. Could saving Group C be justified to Group A? No, because the only way we could prefer Group C is by counting the claims against the mild impairment. This is analogous to the previous case where the single person with a severe impairment cannot accept that we save the group with the mild impairment. After eliminating Group C, we can compare Groups A and B.

This method will hold generally. In effect, it asks us to eliminate the claims that are not relevant to the strongest claim. This means that only those claims that are relevant to the strongest claim can count. Whichever option has the greatest weight of claims that are relevant to the strongest claim, will then be the chosen option.

The justification is coherent and does not render Balance Relevant Claims unattractive. The exclusion of options from the cycle is motivated by the same principle as Balance Relevant Claims itself. It recognizes the primary importance of the person with the strongest claim. Since she has the strongest claim on our help, we need to justify our action primarily to her. Balance Relevant Claims is one way of doing so. Balance Relevant Claims seeks to ensure that our decision to save will always be respectful to everyone involved. The exclusion of options from supposed preference cycles achieves the same goal.

Hybrid Balance Relevant Claims versus Sequential Matching

In their paper “Relevance Rides Again?” Aart van Gils and Patrick Tomlin consider a view similar to the one I have defended.⁴³ They call their view Sequential-Claims Matching. Sequential-Claims Matching is equivalent to what I call Balance Relevant Claims. Unlike my hybrid view, however, there is no additional global relevance condition. For this reason, Sequential-Claims Matching cannot account for the intuition that sometimes small harms, like a sore throat, should not be decisive in decisions about whom to save from severe conditions. At one point in the paper, van Gils and Tomlin consider a condition similar to my global relevance condition. The question they are concerned with is how should we continue to balance after a tie has occurred? The proposal under consideration is that claims can only be counted if they are relevant to the tied claims. For example, in Kamm’s Sore Throat case the sore throat would not be counted because it is not relevant to the claims that constitute the tie, i.e. claims to be saved from death. My global relevance condition can also explain why the sore throat should not be decisive but does this in a different manner. On my view, it is not important whether claims are relevant to the tie but rather whether the decisive claim is relevant to the strongest claim in the group which would not be saved.

The following case which van Gils and Tomlin present against their own proposal illustrates why my global relevance condition is superior.⁴⁴

⁴³ Van Gils and Tomlin, “Relevance Rides Again?”.

⁴⁴ Van Gils and Tomlin, “Relevance Rides Again?,” p. 240. Again, I have changed the precise example to fit it to the stipulations about relevance which I have used in the previous chapters.

Table 4.11: Hybrid Balance Relevance Claims versus Sequential Matching

Group A	Group B
<i>Case One</i>	
1 Death	10 Severe
11 Mild	
<i>Case Two</i>	
1 Death	10 Severe
	1 Moderate
11 Mild	

Following van Gils’s and Tomlin’s proposal, Case One would be a tie. The claims against death and severe impairment are evenly balanced. The claims against the mild impairment are irrelevant to the claims against severe impairment and therefore do not count. In Case Two, however, the claim against moderate impairment is relevant to the tie. This renders the claims against mild impairment relevant to the claim against moderate impairment. Together they outweigh the claim in Group B. We should now save Group A. This means that van Gils’s and Tomlin’s proposal violates the Principle of Addition. The addition of a claim against moderate impairment made Group B less choiceworthy.

The global relevance condition helps avoid this implication. In Case One, the same analysis as before holds for my view. The persons facing severe impairment can veto that claims against mild impairment are counted. The result is a tie. In Case Two, the persons facing severe impairment can again veto the claims against mild impairment. Now the person facing death can veto the claim against moderate impairment. The result is, again, a tie.

The Order of Balancing

Van Gils and Tomlin identify an issue where Balance Relevant Claims seems incomplete. In which order should we balance claims? In the chapter, I started with the strongest claim and then continued to balance less strong claims. The justification for this was that the person with the strongest claim is the one to whom justification is owed most urgently. But priority for the person with the strongest claim could be expressed differently. We could also balance in the interest of the person with the strongest claim. Van Gils and Tomlin call this alternative “Strongest Decides”.⁴⁵ Strongest Decides, they contend, differs from the standard version of Balance Relevant Claims. They illustrate this with the following case.⁴⁶

Table 4.12: The Order of Balancing

Group A	Group B
1 Severe	
11 Moderate	11 Moderate
	11 Mild

Assume we were to balance claims starting with the strongest claim and continuing sequentially but without my global relevance condition. In this case, the claims against mild impairment are decisive. Strongest Decides, on the other side, would balance the claims against moderate impairment with one another. The claims against mild impairment would then be deemed irrelevant to the claim against severe impairment. My Hybrid Balance Relevant Claims can account for the same result. The global relevance condition ensures that the claims against mild impairment cannot be ultimately decisive in not saving the person from severe impairment.

This indicates that Hybrid Balance Relevant Claims can, in some cases, combine starting with the strongest claim and balancing in the interest of the strongest claim. Strongest Decides can differ from standard forms of Balance Relevant Claims by rendering claims irrelevant. In the above example it does so by

⁴⁵ Van Gils and Tomlin, “Relevance Rides Again?,” pp. 250-52.

⁴⁶ Van Gils and Tomlin, “Relevance Rides Again?,” p. 250. Once more I adjusted the relevance stipulations for sake of consistency with the ones I have used.

creating a “gap” in the line of claims to be considered. It balances claims of equal strength with one another and thereby leaves some claims in a group to be irrelevant to the overall strongest claim. This is also what my global relevance condition does. If claims that are irrelevant to the strongest claim would tip the balance, they become irrelevant. My view does not need to alter the order in which claims are balanced to account for this.

Strongest Decides can also differ from other forms of Balance Relevant Claims, including mine, in a different manner. Consider the following case.

Table 4.13: Hybrid Balance Relevance Claims versus Strongest Decides

Group A	Group B
(1 Severe)	(10 Moderate) + (1 Moderate)
(10 Mild) + (100 Mild)	(100 Mild) + 10 Mild

In this case, my view says that we ought to save Group B. If we were to not count the claims of mild impairment, the person with severe impairment would get counterbalanced by the claims against moderate impairment itself. In a case like this it is possible for Strongest Decides to differ from my view. It would approach the case as follows.

Table 4.14: Hybrid Balance Relevance Claims versus Strongest Decides (II)

Group A	Group B
1 Severe	(11 Moderate)
(110 Mild)	110 Mild

By balancing the claims against moderate impairment in Group B with the claims against mild impairment in Group A, Strongest Decides renders the claims against mild impairment in Group B irrelevant. My global relevance condition would not have deemed the claims against mild impairment irrelevant. This is because doing so would also render Group A’s claims against mild impairment irrelevant. But the person with the severe impairment needs these claims to count against B’s claims against moderate impairment. This means that Strongest Decides differs from my approach here. It also highlights the problem with Strongest Decides, namely it

violates *Equal Consideration for Equal Claims*. Claims of mild impairment are considered and balanced if they are in Group A but deemed irrelevant if they are in Group B.

Van Gils and Tomlin discuss the violation of *Equal Consideration for Equal Claims* and suggest the following argument in defense of Strongest Decides.⁴⁷ Strongest Decides does not violate a narrower version of *Equal Consideration for Equal Claims* according to which it is impermissible that some claims of equal strength are disregarded at the outset while others are not. In my example it is not the case that some claims against mild impairment are disregarded at the outset. It is only because there are multiple ways of balancing claims against one another and the fact that the strongest claim chose this particular way of balancing. I remain unconvinced that this form of violating *Equal Consideration for Equal Claims* is meaningfully different. The core idea of *Equal Consideration for Equal Claims* is that claims of equal strength should have equal moral force. If we consider some claims and include them in our argument for why we save a particular group, we cannot then also say that claims of equal strength are irrelevant. If the claims against mild impairment of a member of B are irrelevant, then why is it fine to consider the claims against mild impairment of a member of A as relevant in deciding that we should save A? Any member of B facing a mild impairment can rightfully complain that we did not accord her with the same moral concern as the member of A.

Thus, I believe that we should accept at most that we balance in the interest of the strongest claim unless doing so violates *Equal Consideration for Equal Claims*. This means that we rule out this second case. We are left only with the first case. But, as I have argued, in this case my global relevance condition already ensures that a form of balancing that starts at the top is also in the interest of the person with the strongest claim. Strongest Decides is not an alternative to my view.

⁴⁷ Van Gils and Tomlin, "Relevance Rides Again?," pp. 251-52.

Chapter 6. Constraints, You, and Your Victims

*“Individuals have rights, and there are things
no person or group may do to them”.*

Nozick. *Anarchy, State, and Utopia*. p. ix

I. Introduction

The quote with which Robert Nozick begins his *Anarchy, State, and Utopia* captures a position widely shared among deontologists. Nozick maintains that rights are not only one important normative consideration among others, but that they exclude some options available to us. Someone else’s right to life means that it is wrong for us to murder a person, full stop. Someone else’s right to privacy means that it is wrong for us to invade a person’s privacy, full stop. Nozick calls this position “rights as side constraints”.¹ Following this view, it is impermissible to violate rights even if we thereby prevent more violations of the same right from happening. Assume I could save five people from being murdered by killing one person. Rights as side constraints would condemn the act, because our morally permissible options are constrained by the violation of the single person’s right. An action can be morally permissible only if it respects all side constraints.²

Not all deontologists would agree with Nozick’s understanding of rights. Indeed, many deontologists think that rights can be weighed against one another and overridden. Rights enjoy a certain normative priority over other moral considerations, but they do not have the power to categorically exclude courses of action. Deontologists often capture this status of rights by introducing a distinction between infringed and violated rights.³ In this chapter I do not want to enter the

¹ Nozick, *Anarchy, State, and Utopia*, p. 29.

² Nozick leaves open the question whether there may be exceptions to side constraints in the case of “catastrophic moral horror”. Nozick, *Anarchy, State, and Utopia*, p. 29fn.

³ Joel Feinberg, *Rights, Justice, and the Bounds of Liberty* (Princeton: Princeton University Press, 1980), pp. 229-32; and Thomson, *The Realm of Rights*, pp. 82-104.

discussion among deontologists whether or not side constraints are co-extensive with rights, because in spite of the different positions on rights, deontologists share a common commitment. This commitment holds that there are at least some constraints on our actions that function in the way Nozick suggests.

The view that morally permissible actions have to respect all side constraints C is subject to a powerful criticism. Nozick appears to be the first to have articulated this criticism against his own view. Given that we care about the moral values that ground the side constraints C, isn't it irrational to refuse to prevent as many violations as possible?⁴ How can one square caring about the non-violation of C while standing idly by when one could prevent many of these violations? Adherence to side constraints appears almost like a rights fetishism. Under a rival view, which Nozick calls "utilitarianism of rights" we ought to minimize rights violations. This view would still not be a form of utilitarianism as long as we determine which rights individuals have in a non-utilitarian manner. As such, utilitarianism of rights escapes some of the standard criticisms standard utilitarianism faces.⁵

There are (at least) two ways of responding to the irrationality objection. The first approach is an agent-based approach. It focuses on the agent making the decision and seeks to find something special about either her or the relationship she has with the victim. The second approach is a victim-based approach. This route to justifying constraints seeks to find a special feature in the victims and would-be victims. One way to bring out the contrast between agent- and victim-based justifications is the following. An agent-based justification focuses on the agent and what makes it impermissible *for her to kill*. A victim-based justification focuses on the victim and what makes it impermissible *to kill him*.⁶

⁴ Nozick, *Anarchy, State, and Utopia*, p. 30. The objection is sometimes also known as the "paradox of deontology", a terminology due to Samuel Scheffler. See Scheffler, *The Rejection of Consequentialism*, pp. 80-114; and Samuel Scheffler, "Agent-Centered Restrictions, Rationality, and the Virtues," *Mind* 94 (1985): 409-19.

⁵ If, however, we determine rights instrumentally in virtue of the well-being that having these rights brings about, then utilitarianism of rights turns into a version of rule utilitarianism. It might then be described as a "negative rule utilitarianism" combining the focus on justifying general rules for actions together with a focus on minimizing evil instead of promoting good.

⁶ Some defenses of deontological constraints defend constraints on the basis of a rejection of a maximizing conception of rationality. The irrationality objection dissipates, these authors argue, because rationality does not require us to promote all values. (See Paul Hurley, "Agent-

While most accounts of side constraints are victim-based and many are based on the idea of inviolability, my aim in this chapter is to outline and defend an alternative, agent-based justification.⁷ We do not need to appeal to inviolability in order to justify deontological constraints. By focusing on the contribution of individual agents and their actions as opposed to the actions of *different* agents, this justification illustrates the importance of the separateness of agents that I highlighted in the introduction.⁸ Frances Kamm classifies agent-based justifications into three different groups. They can be *agent-relative* by giving different basic aims to different agents. They can be *agent-focused* by emphasizing a quality of agency. They can be

Centered Restrictions: Clearing the Air of Paradox," *Ethics* 108 (1997): 120-46; and Scanlon, *What We Owe to Each Other*, pp. 81-86.) Such a defense is incomplete, however, unless it gives us an account of why respect for rights should take the form of deontological constraints. Most would agree that we should save the greater rather than lesser number. Why not then violate one right to save many from rights violations? Answers to this question can again focus on the agent or on the victim. A genuinely third way is argued for by Christopher McMahan who argues that because there are other ways how to prevent the badness of rights violations we are not forced to minimize rights violations (Christopher McMahan, "The Paradox of Deontology," *Philosophy & Public Affairs* 20 (1991): 350-77.) Constraints are not as solidly founded under this picture given that minimizing rights violations is still an available option to agents as McMahan admits.

⁷ The standard account is Frances Kamm's (F.M. Kamm, *Morality, Mortality*, vol. 2 (New York: Oxford University Press, 1996), pp. 259-89; Kamm, *Intricate Ethics*, pp. 17-31, 130-89, 227-84). Nozick's justification resembles Kamm's in many ways (*Anarchy, State, and Utopia*, pp. 30-50). Other prominent accounts that rely on inviolability are Warren Quinn's ("Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing," *Philosophical Review* 98 (1989): 287-312, at pp. 305-12) and Thomas Nagel's ("Personal Rights and Public Space," *Philosophy & Public Affairs* 24 (1995): 83-107, at pp. 89-99). Kasper Lippert-Rasmussen defends a revision of Kamm's account that is also based on the idea of moral status ("Moral Status and the Impermissibility of Minimizing Violations," *Philosophy & Public Affairs* 25 (1996): 333-51). Richard Brook gives a victim-based justification that does not rely on inviolability or moral status ("Agency and Morality," *Journal of Philosophy* 88 (1991): 190-212, at pp. 201-9). For critical discussion see Shelly Kagan, "Replies to my Critics," *Philosophy and Phenomenological Research* 51 (1991): 919-28, at pp. 919-22; Lippert-Rasmussen, "Moral Status and the Impermissibility of Minimizing Violations"; David McNaughton and Piers Rawling, "On Defending Deontology," *Ratio* 11 (1998): 37-54, at pp. 48-53; Michael Otsuka, "Are deontological constraints irrational?" in *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, ed. Ralf M. Bader and John Meadowcroft (Cambridge: Cambridge University Press, 2011), pp. 38-58; and Susanne Burri, "Personal Sovereignty and Our Moral Rights to Non-Interference," *Journal of Applied Philosophy* 34 (2017): 621-34.

⁸ I thereby respond to Frances Kamm who has rejected the idea that the separateness of agents can ground deontological constraints. See Kamm, "Moral Status and Personal Identity," pp. 290-91.

agent-concerned by focusing on what the violation of the constraint does to the agent.⁹ My justification is not any of the three above but rather a *relational agent-based* justification. I focus on the relation between the agent and the victim.¹⁰

After outlining my relational agent-based justification, I will, in Section III, introduce the problem of minimizing one's own rights violations. Accounting for the wrongness of violating one right in order to prevent oneself from violating multiple rights poses the strongest challenge to agent-based justifications and I rise to this challenge in Sections III and IV. Before concluding, I will respond to two further criticisms of agent-based justifications, namely the charge that agent-based justifications are self-indulgent (Section V) and that they imply an unappealing symmetry between cases involving persons and non-persons (Section VI).

II. A Relational Agent-Based Justification for Side Constraints

The irrationality objection compares two different states of affairs. To illustrate: A sadist has pushed a trolley towards five persons who will die if the trolley hits them. The only way for you to prevent this is by pushing an innocent bystander in front of the trolley, thereby killing her and stopping the trolley. In state of affairs A, you do nothing and five persons' right to life is violated. In state of affairs B, you kill the bystander and one person's right to life is violated. Surely, the violation of five rights is worse than the violation of one right, it is argued. However, there is, of course, the following relevant difference. In the second case, it is *you* who is violating the right, in the first case it is *someone else*.

⁹ Kamm, *Morality, Mortality*, 2:238. A standard formulation of an agent-relative justification is given by Samuel Scheffler who ends up rejecting this argument (Scheffler, "Agent-Centered Restrictions, Rationality, and the Virtues"). For a defense see McNaughton and Rawlings, "On Defending Deontology". Thomas Nagel provides a justification (*The View From Nowhere*, pp. 175-85) which can be seen as an example of an agent-focused approach. Stephen Darwall's justification can be interpreted as an example of the agent-concerned approach ("Agent-Centered Restrictions from the Inside Out," *Philosophical Studies* 50 (1986): 291-319).

¹⁰ For a different example of the relational approach see Alec Walen, "Doing, Allowing, and Disabling: Some Principles Governing Deontological Restrictions," *Philosophical Studies* 80 (1995): 183-215, at pp. 185-90. Unlike my argument, Walen does not make clear how his justification avoids the standard criticisms which agent-based justifications face.

We have a special responsibility for our own actions, a responsibility that is greater than the responsibility we have for actions we let happen. To sharpen our understanding of this difference, consider the following pair of cases.¹¹ Late at night you are driving through a scarcely populated area. You see at the side of the road a person who is bleeding and badly injured. There is no risk for her life, but she is in great pain. Given that you are morally motivated you decide to help her. Your phone does not have coverage, so you need to drive her to the nearest hospital. On the way your car breaks down. Luckily you see lights in a house nearby and decide to ask for help. An elderly lady opens the door but upon the sight of your bloody hands she locks herself into a room. The house does not have a telephone but only a car which you could use to help the injured stranger. The keys are unfortunately in the room where the old lady is locked in. In your desperation you look for solutions to get her to leave the room. The only solution you come up with is to use her grandchild which she left behind in the rush and twist her arm. Hearing her grandchild scream you are certain she will leave the room. Still, here you should not use the innocent child as a means to get the car keys.

Contrast this with a case where you drive by the house and see the lady is about to twist the child's arm. You have the chance to stop bringing the stranger to the hospital by stopping and intervening in the fight.¹² But here, we think, you are not required to do so. The pain suffered by the stranger is by far greater than what the child will, unfortunately, experience. How do we explain this asymmetry? The easiest way is by appealing, as I already did, to the special responsibility we have for our own deeds. It would be *you* who hurts the child in the first case while it would be someone else in the second case. What is remarkable about these cases is that here facts about our own agency can overcome judgments about the wrongness of the different actions. This shows that facts about agency have a deeper role than merely being tie-breakers.

¹¹ The cases closely follow Nagel, *The View From Nowhere*, pp. 176-77.

¹² Assume it would take equally long to protect the child from the assault than it would to get the car started again in the earlier case.

There have been different attempts to explain why precisely we have this special responsibility. Thomas Nagel focuses on one aspect of our agency.¹³ One morally central part of our agency are intentional actions. What is special about our intentions is that our actions are guided by them. We adopt specific aims. In cases where we can do evil to prevent even more evil it would mean that we intend evil, we let ourselves be guided by evil aims. But, as Nagel puts it, “the essence of evil is that it should *repel* us”.¹⁴ This explains why evil should not be brought about intentionally by us. But what about evil that is allowed because of an evil intention? You might not stop to help the child not because you want to bring the stranger to the hospital, but because of your hatred of children. You let the harm happen out of an evil intention. It is even possible that your inaction (not helping) may be guided by evil aims.

We can refine Nagel’s point with an argument made in defense of the doctrine of double effect. What counts is not whether or not agents have evil intentions. What counts is whether they have a justification for their action that does not require evil intentions.¹⁵ Evil should repel us at all times, but the repelling evil makes an act impermissible only when the repelling evil is unavoidable. We can illustrate this by considering our reasons for action. If the child were to object to our twisting her arm and voice her pain, this would only mean that we have succeeded. Only if the child is in pain we have achieved our aim. The child’s objection constitutes a reason for our action. This is not the case when we do not intervene and let harm happen. The harm that accrues to the child if the old lady twists the child’s arm does not have to be a reason for our inaction. We can wish that the child was not hurt.

An alternative, and complementary, strategy departs from the idea that each of us has a special responsibility for one’s own life. We each have our own life to lead and it is incumbent upon us to lead a good life. The strategy is then to argue that the special responsibility for our lives should lead us to accept the special responsibility for our own actions. To see this, consider how we should treat our own actions if we

¹³ Nagel, *The View From Nowhere*, pp. 180-85.

¹⁴ Nagel, *The View From Nowhere*, p. 182.

¹⁵ See William J. FitzPatrick, “Acts, intentions, and moral permissibility: in defence of the doctrine of double effect,” *Analysis* 63 (2003): 317-21.

did not have any special responsibility for them. We should take a purely instrumental view on our actions. What would matter is the overall outcome in the grand causal web, not what we contributed to it as opposed to others. It would not matter what the contribution of specific actions and specific persons are. The only thing that would matter is whether the opportunity to act (or fail to act) resulted in the optimal outcome. Such an instrumental attitude towards our actions cannot make sense of feelings of remorse or regret for what you in particular have done. The instrumental attitude can accept these sentiments only as irrational feelings that are a bad thing that happens. But this is not the content of these sentiments. Having done harm is not a bad thing that happened to a person. It is a flaw in that person's life, a stain on one's moral record. A person's character, sense of self and integrity are bound up with what this person does. We identify with how we act. This identification and sense of integrity are part of our moral agency.¹⁶

Just as we need to honor our own special responsibility for our own life, we need to respect the special responsibility of others for their lives. Doing harm means infringing in the sphere of control of someone else. It usurps the decision how a person's life and body are to be used and thereby amounts to commandeering someone else's life. This denies the other person her special responsibility for her life. Allowing harm, on the other hand, does not involve such commandeering of another person's life. It does not involve a decision how this person's life and body are to be used.¹⁷ Ronald Dworkin illustrates this idea with a swimming metaphor. The responsibility for our own life means that we are like swimmers who swim in different lanes. We are allowed to concentrate on swimming in our lanes and only sometimes are required to cross lanes in order to aid others. But what is more strictly forbidden is the crossing of lanes and the interference with others.¹⁸ This indicates

¹⁶ See Williams, "A Critique of Utilitarianism," pp. 93-118.

¹⁷ Quinn, "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing," pp. 308-10; Kamm, *Intricate Ethics*, pp. 17-21; and Dworkin, *Justice for Hedgehogs*, ch. 13. This explanation may also explain why harming by merely foreseen but unintended side effects can be permissible. I discuss this issue shortly when discussing the limits of my construal of the distinction between killing and letting die.

¹⁸ Dworkin, *Justice for Hedgehogs*, pp. 287-88.

that the special responsibility for our lives justifies the special responsibility for our actions.

The special responsibility for our own actions and choices allows us to view the moral dilemma in the case of the sadist I mentioned at the beginning of this section from a first-personal perspective. Our special responsibility just means that there is something different about this perspective. Now from this perspective the action appears very different than from the impersonal perspective that compares states of affairs. From your first-personal perspective, you are confronted with your relation to the six individuals in the situation. What complaints can they make against your actions? The one who would be killed by you to save the five can complain that you would kill him. This is just like the girl in my earlier example who could complain that you would twist her arm and hurt her. Now the five others also have a complaint not to be killed. But this complaint is not directed to you, but to the original perpetrator causing the entire dilemma. To you, their complaint is still serious, but it is less serious than to the original perpetrator. If you would not discount the complaint to you, this would mean your relation to the victim was the same as the relation of the original perpetrator. This seems wrong. The parallel is here with the girl in the second case. Her complaint not to be hurt by having her arm twisted is directed to the old lady. To you she can only complain that you did not save her. In cases involving the killing of one you are thus facing one complaint against being killed versus five complaints against a failure to be saved. Given that we discounted the complaint of the five, it seems now plausible that you should act on the more serious complaint of the one. This is what we mean when we say that killing one is worse than letting five die.

My construal of the distinction between killing and letting die can also explain the limits of this distinction. For example, it seems to many that it is permissible to save five people by diverting a trolley to a sidetrack where it would kill one bystander as an unintended side effect. In this case it would then be permissible to kill the one instead of letting the five die. Nagel's understanding of the special responsibility for our actions can help explain this. It is a crucial component of Nagel's view that intending evil is what makes acts impermissible. Insofar as the death of the bystander

is not intended, the case would not be covered by Nagel's explanation. Similarly, Dworkin argues that using a person in order to save others constitutes usurpation of control. Insofar as the death of the bystander is not necessary for the saving of the five, we have not used the one in order to save the five. The trolley case would then not be covered by Dworkin's explanation either. My argument in this article is compatible with both the possibility that diverting the trolley is permissible and with the possibility that it is impermissible.¹⁹ To avoid misunderstandings, any reference to the distinction between killing and letting die should therefore be understood as a difference between intentional killing or killing as a means or end and letting die.²⁰ A second limit includes such killings that are not wrongful such as killing in self-defense. Viewed from your first-personal perspective killing responsible aggressors in self-defense would not create any problem. If we take seriously the idea that responsible aggressors have made themselves liable to be killed, then a responsible aggressor would not have any grounds to complain to you. Therefore, we should understand further references to the distinction between letting die and killing as between letting die and pro tanto wrongful killing.²¹

Deontological constraints can therefore be understood in terms of the relational construal of the distinction between killing and letting die. The reason why you are not permitted to violate the right of the one in the case of the sadist is because of the stronger complaint that the would-be victim has against you. Killing her would place you in a different relation to her than the relation in which you stand to the five that you let die.

¹⁹ For the bystander case see Judith Jarvis Thomson, "The Trolley Problem," *Yale Law Journal* 94 (1985): 1395-1415, at pp. 1396-99. Thomson has later expressed some doubt on the permissibility of turning without ultimately endorsing either position. Judith Jarvis Thomson, "Turning the Trolley," *Philosophy & Public Affairs* 36 (2008): 359-74.

²⁰ Whichever of the two depends on what one takes to be the salient explanation for why diverting the trolley in cases like the bystander case is (or seems) permissible. Intentional killing indicates a preference for the Doctrine of Double Effect, killing as a means indicates a preference for the Means Principle.

²¹ One further comment: I adopt the phrase "killing is worse than letting die" and variations of it, given its familiarity. I think the phrase is misleading in some sense. The issue is not about whether killing or letting die is axiologically better, but rather whether one or the other is more justifiable. "It is harder to justify killing than to justify letting die" would be a more accurate statement.

III. The Puzzle of Minimizing One's Own Rights Violations

A. *One's Own Violations and the Guilty Agent*

There is a powerful objection to the view and the reasoning that I just articulated. Kamm expressed this objection with her *Guilty Agent* case. You have set a bomb which will kill five people. Later you have a moral epiphany and realize that you ought not to have done this. The only way to prevent the bomb from killing the five is by shooting a sixth person and placing her body over the bomb.²² In this case my reasoning would seemingly lead to the conclusion that you ought to kill the one. After all, the complaints all six have towards you seem identical. Everyone can now complain to you that you would be killing her. More generally speaking, the case has the interesting twist that here you would prevent *yourself* from killing five people by killing one. It seems that you now have the option between (intentionally) killing five and (intentionally) killing one. But at the same time the deontological constraint against killing does not suddenly disappear because of some previous wrong you committed in the past.

This objection is indeed defeating for one version of an agent-based justification for constraints. We might conceive our special responsibility for our own actions as meaning that we should minimize our own violations of constraints. Kamm calls this justification *agent-relative*.²³ Under this reconstruction of agent-based justifications we have not abandoned the idea of maximization altogether. We are still committed to the requirement of rationality to maximize our desired outcome. But the facts about our special responsibility of our own actions and our own agency make a difference nevertheless. They lead us to consider maximization relative to

²² Kamm, *Morality, Mortality*, 2:242; and *Intricate Ethics*, pp. 26-27. Kamm credits Alan Zaitchik for coming up with cases of this kind. See Alan Zaitchik, "Trammell on Positive and Negative Duties," *The Personalist* 58 (1977): 93-96. Cases of this kind are also discussed by Judith Jarvis Thomson and by Richard Brook. See Thomson, "The Trolley Problem," pp. 1399-1401; Thomson, *The Realm of Rights*, pp. 139-140; and Brook, "Agency and Morality", pp. 197-99.

²³ See McNaughton and Rawling, "On Defending Deontology"; Ulrike Heuer, "The Paradox of Deontology, Revisited," in *Oxford Studies in Normative Ethics. Volume 1*, ed. Mark Timmons (Oxford: Oxford University Press, 2011), pp. 237-67; and Christa M. Johnson, "The Intrapersonal Paradox of Deontology," *Journal of Moral Philosophy* 16 (2019): 279-301.

each agent. We act in accordance with such an agent-relative moral goal by abstaining from acting in the usual dilemma case. We minimize our own killings by not killing anyone but allowing the killing of the greater number. The *Guilty Agent* case shows that this solution does not work. There are cases in which we may be forced to violate a constraint in order to minimize our own violations.²⁴

One response to the *Guilty Agent* would be to narrow down the scope of the minimizing requirement further. In addition to being agent-relative, the goal would also be temporally-relative. While the principle at stake here has the intuitively right answers, it is weak in justifying them. Why should we attach such great moral significance to the temporal specification? Perhaps the difference between now and a year later is seen as significant. You poison five people now and in a year time you frantically try to save them from the seemingly inevitable death. The only way to do so is by killing one person whose organs you can use to brew an effective antidote. Maybe we think that this time difference is significant. But what about now and tomorrow? Now and in one hour? Now and a moment from now? The latest difference does not carry any moral significance, yet the permissibility does not change depending on the time interval.²⁵

Christa Johnson, in a recent defense of constraints that is both agent-relative and time-relative, attempts to provide an answer to the question why differences in time are morally important. In order to justify time-relativity, Johnson invokes the “appeal to full relativity” according to which all reasons that speak in favor of agent-relativity also speak in favor of time-relativity.²⁶ Yet it is hard to see why this is the case. I argued that the best support for agent-relativity comes from the distinction between killing and letting die and the special responsibility we have for our actions. Thinking about one’s own responsibility as an agent would, if anything, seem to

²⁴ Heuer (“The Paradox of Deontology, Revisited”) bites the bullet and argues that deontological constraints do not in fact apply to this kind of case. Along with the vast majority of authors writing on this subject I will assume that deontological constraints should apply to the problem of minimizing one’s own violation.

²⁵ See Brook, “Agency and Morality,” pp. 198-201; Kamm, *Intricate Ethics*, p. 27; Otsuka, “Are deontological constraints irrational?,” pp. 44-46.

²⁶ Johnson, “The Intrapersonal Paradox of Deontology,” pp. 291-292; see also Nagel, *The Possibility of Altruism*, pp. 16-19, 99-100; and Parfit, *Reasons and Persons*, pp. 137-148.

speaking in favor of time-neutrality. After all, one is equally responsible for all parts of one's life.²⁷ Secondly, there seems to be an important difference between different persons and different points in time. The separateness of persons has moral significance in a way that the separateness of time points has not. This explains a clear asymmetry between agent-relativity where the separateness between one's own life and the lives of others matters and time-relativity for which no comparable argument can be made.²⁸

B. *Responding to the Guilty Agent*

When I introduced the *Guilty Agent* case I said that it seems that you are facing the choice between killing five and killing one. Therefore, we cannot appeal to the distinction between doing and allowing in order to justify why you are facing a constraint against killing in this case as well. However, I need to correct myself. I do not think that this interpretation of the case is the correct one. Indeed, it is this mistaken interpretation which makes the *Guilty Agent* appear to be such a powerful objection to agent-based justifications. Once we see this interpretation is inaccurate, agent-based justifications will become a much more viable option than before.

My argument proceeds in two steps. In the first step I argue that the choice the *Guilty Agent* is facing is one between doing and allowing. The argument is that the best description of the agent's choices includes one of doing harm and the other of allowing harm. In the second step I examine the suggestion that the *Guilty Agent*'s allowing harm is relevantly different from standard cases of allowing harm. There are two possibly relevant disanalogies. First, the *Guilty Agent* is allowing the victim to be killed rather than merely allowing her to die. Second, the *Guilty Agent* is allowing the victim to be killed *by herself* rather than being killed simpliciter. However, I reject the suggestion that these forms of allowing harm are relevantly different from the standard case of allowing harm in the *Guilty Agent* case.

²⁷ See also Brook, "Agency and Morality," p. 199.

²⁸ Brink, "Rational Egoism and the Separateness of Persons".

My response may seem similar to invoking time-relativity which I rejected earlier. However, there are two important differences. First, unlike the agent- and time-relative view that I criticized, my justification does not rely on assigning agents different agent-relative goals. Instead, my justification highlighted that the objection to minimizing rights violations is grounded in the way the agent would relate to her victims if she were to kill the one to save the five from being killed. Second, I argue that the distinction between different actions is morally significant rather than the distinction between different points in time. Because different actions are carried out after one another it may seem that the difference in time is relevant. However, I argue that any significance of time is only derivative of the significance of distinguishing between separate actions. Thereby, I provide a more satisfactory answer to the question why differences in time can be morally important.

C. *First Step: Choice between Doing and Allowing*

To support my claim that in the situation of the *Guilty Agent* you are still facing a decision of doing versus allowing, consider my *Inconclusive Agent*. An agent has stabbed a person and wounded her seriously. The victim is suffering from severe blood loss, but not yet dead. It will take, say, half an hour until she dies from the blood loss. The agent is still present at the scene, she takes a deep breath and then starts to contemplate whether or not she should deliver first aid. She is inconclusive. On one side she is a sadist and enjoys seeing her victim suffering, on the other side she has some appreciation of the moral demand not to kill people. While she is contemplating, what is the agent doing? It seems that the agent is allowing the death of her victim by not delivering first aid. Her first action, the stabbing, is over and now she is engaged in another, different action. What makes this case peculiar is that the results of the first action are not yet known. It may be that the stabbing amounts to killing the victim (if she does not deliver first aid), it may be that it does not (if she delivers first aid). Therefore, we cannot fully describe what the first action was. It may be that the stabbing is best described as a killing or rather as merely attempting to kill.

My analysis of the *Inconclusive Agent* is strengthened by what Thomson calls the Reductive Theory of Action.²⁹ According to the Reductive Theory of Action, every action is a set of bodily movements caused by intentions in the right way. The fact that an act can be given different descriptions only indicates that the same act can be described differently. The intention to kill caused the bodily movements which constituted the stabbing. When contemplating whether or not to continue, the agent is engaged in a different set of bodily movements or lack thereof.

Even though my analysis follows naturally from the Reductive Theory of Action, a similar argument can be made assuming a rival view of action according to which differently described acts are numerically distinct acts.³⁰ This view has two notable features. First, acts which the Reductive Theory of Action classifies as numerically identical stand in a relation of “amounting to”. The agent’s stabbing amounts to killing, or put otherwise, the agent kills *by* stabbing. Second, these actions can have different temporal extensions. In the *Inconclusive Agent*, the action of killing is continuing until the death of the victim. Nevertheless, this does not mean that the subsequent deliberation of the agent is part of the killing. The *Inconclusive Agent*’s stabbing rather than her contemplating will amount to the killing, and the stabbing is over. Even on this view, part of the killing, namely the stabbing, is in the past. When contemplating between delivering first aid or not, killing the victim is not one of the options she can choose from. She can only choose future actions. She can choose between waiting and helping, for example.³¹ These future actions may, in a similar spirit to the Reductive Theory of Action, then determine whether there will have been

²⁹ Thomson, *The Realm of Rights*, pp. 125-27. For a proponent see Donald Davidson, “Actions, Reasons, and Causes,” *Journal of Philosophy* 60 (1963): 685-700, at pp. 686-87; and Donald Davidson, “The Individuation of Events,” in *Essays in Honor of Carl G. Hempel*, ed. Nicholas Rescher (Dordrecht: Reidel, 1969), pp. 295-309. Other philosophers like G.E.M. Anscombe or Jonathan Bennett concur with the Reductive Theory of Action that co-occurring actions like the killing and stabbing are one and the same action, differently described. See G.E.M. Anscombe, *Intention*, 2nd edn. (Cambridge, MA.: Harvard University Press, 1963), pp. 11-12, 37-47; and Jonathan Bennett, *Events and their Names* (Oxford: Oxford University Press, 1988), pp. 188-202.

³⁰ Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs: Prentice-Hall Inc., 1970), chs. 1-3; Lawrence Davis, “Individuation of Action,” *Journal of Philosophy* 67 (1970): 520-30; and Judith Jarvis Thomson, “The Time of a Killing,” *Journal of Philosophy* 68 (1971): 115-32.

³¹ This is supported by Judith Thomson, a proponent of this theory of action. Thomson, “The Trolley Problem,” pp. 1399-1400; Thomson, *The Realm of Rights*, pp. 139-140.

a killing.³² But this does not make “killing” one of the options the Inconclusive Agent can choose when contemplating, since her waiting will not amount to a killing.

The phenomenon that future actions can influence past actions is common for many actions whose typical outcomes we only see in the future. We can, of course, describe the action more basically as a stabbing.³³ But the important description here is one that attaches special meaning to the action, like killing or letting die. To take a non-moral example, imagine a football player who shoots the ball aiming at goal. We do not know immediately whether the player scored a goal. This will depend on what other agents, in that case the goalkeeper, do. The standard case is this one where the description of our action depends on what *other* agents do. In the *Inconclusive Agent* case the interesting factor is that here the agent herself can change the description of the action after the action occurred. Should this make a difference for describing what the agent is doing in her second action? Should this turn an allowing into a doing? I cannot see why it should. For imagine that we learned that the Inconclusive Agent was not the agent who stabbed the victim. But in all other respects the Inconclusive Agent does the very same thing as before. Why should this turn her act into a doing as opposed to an allowing? She, like the perpetrator, sits next to the victim and goes through the very same thought processes while refraining from doing anything. It would be artificial to draw a line between the two agents and say that one is currently doing harm while the other is allowing it.

Further, is this case relevantly different from the *Guilty Agent* case? It does not seem so. In the *Guilty Agent* case the first action, setting the bomb, is over. The Guilty Agent is, like the Inconclusive Agent, waiting for her action to yield results. Then, the Guilty Agent sees the opportunity to engage in a different action when she spots the sixth man who could mitigate the detonation. The Inconclusive Agent similarly sees

³² Following this non-reductive theory of action, it can then be indeterminate whether an action is going on or not because this will depend on future events. Only if the Inconclusive Agent fails to intervene, it will be true that she was killing all along. This is one of the reasons why I reject the non-reductive theory and follow the reductive theory in my formulations. Another reason is that it makes it possible for acts to continue after the death of the agent. If the stabbing causes the victim to die after a prolonged coma, the killing would continue until the death of the victim. But it is possible for the perpetrator to die before the victim dies. If so, the current theory implausibly holds that the act of the agent continues after her death.

³³ Or focus on the more basic action of stabbing if we reject the Reductive Theory of Action.

the opportunity to engage in a different action when she spots the first aid kit which could save her victim's life. If the *Inconclusive Agent* refrains from acting, she lets the person be killed. If the *Guilty Agent* refrains from acting, she similarly lets the five people be killed. However, should the *Guilty Agent* decide to shoot the one, she would be killing the one person. The decision therefore is not one of killing one or killing five, the decision is one of killing one or allowing five to die.

D. Second Step: Killing versus Letting Be Killed

One worry about my argument stems from the idea that not all forms of letting die are equal. The claim that killing is worse than letting die is familiar. But in the *Guilty Agent* and *Inconclusive Agent* cases you are facing a choice between killing and letting be killed. Perhaps this makes a difference.

I submit that it does not.³⁴ To see this, consider the following case. You stand in the middle of two tracks. Two trolleys are approaching, one on each track. On both tracks there are workmen whom you try to warn but who are unable to hear or see you. There is a lever that you can pull which would divert one of the two trolleys onto an empty sidetrack. You also know the following. The trolley on the left was set in motion by a villain, the trolley on the right broke loose naturally. Should you have any preference whom to save? In the case of the right-side trolley, the deaths would be an unfortunate accident. The workmen would not be killed in a rights-violating manner. In the case of the left-side trolley the workmen would be killed by the villain. If there is a difference between letting die and letting be killed, then we ought to save the workmen from the left-side trolley. Yet it would certainly not be wrong to save the workmen from the right-side trolley or to flip a fair coin to determine who should be saved. It would be permissible to save either of the two groups or to give equal chances to both.

³⁴ For similar arguments and sentiments see Scheffler, *The Rejection of Consequentialism*, pp. 109-110; Nagel, *The View From Nowhere*, p. 178; Thomson, *The Realm of Rights*, pp. 137-39, 142-43; Scanlon, *What We Owe to Each Other*, p. 83.

E. *Second Step: Killing versus Letting Be Killed by Oneself*

There is something peculiar, however, about both the *Inconclusive* and the *Guilty Agent* that is absent in my villain trolley case. In both cases we can add information having to do with the particular causal history of the agents facing the later choice. Not only do the two agents let their victims die, but they also let *themselves* become the killer of their victims. Now should this make a difference? We accepted that killing is worse than letting die. We also accepted that killing is worse than letting be killed. Should we think that letting oneself become the killer is worse than letting die (or be killed)?

The best way to understand the proposal of “letting oneself become the killer” relates to the special normative situation in which we are once we have committed wrongs.³⁵ Here it seems plausible that we have special obligations towards those we have wronged to make up for our wrong. Letting oneself become the killer carries some moral weight over and above letting die (or be killed). The reason here is similar to other special obligations. Consider the case of family bonds. Letting one’s partner die (or be killed) carries moral weight over and above letting die (or be killed).

The context in which special obligations are most impactful are acts that would be discretionary if it was not for the presence of special obligations. For example, you are not required to drive a stranger to the hospital so that she receives care for her sprained ankle. You may do so, but if you do, your act would be supererogatory. The presence of a special obligation changes this picture. You are

³⁵ This is also suggested by Jason Hanna in an article on the difference between doing, allowing, and allowing one’s own doing. Hanna goes on to argue that a problem with this approach is that it cannot account for cases where current actions can prevent future actions given that in those cases the agent has not yet committed any wrong (Jason Hanna, “Doing, Allowing, and the Moral Relevance of the Past,” *Journal of Moral Philosophy* 12 (2015): 677-98, at pp. 680-89). In order for the problem to get off the ground Hanna needs cases in which it is certain that harm will be done. This cannot be said if the future action is one that is under the control of the agent. An agent could not claim that harm now was necessary to prevent future harm if the agent could have chosen not to harm at a later stage. The cases therefore involve harm that is caused in the absence of agency, for example during sleepwalking or due to mental incapacitation. I do not believe that a unified explanation for both of these cases is needed. Cases that involve killing without agency introduce further complications. These cases also do not raise the specific problem of the irrationality objection since it is doubtful whether such killings are rights violations.

required to drive your partner to the hospital for the sprained ankle treatment. In a similar vein, you would be required to drive a stranger to the hospital if you had wrongly sprained their ankle. (Assuming the stranger would be willing to let you drive her to the hospital.) Many other typical obligations towards one's victims are part of this category. The obligation to apologize for our wrong is one example, as well as the obligation to compensate. There is little that speaks against these obligations here since no further person is entitled to your actions.

Special obligations can also have an effect in a different context. Consider situations in which you can save only some but not everyone from harm. For example, many people are injured and need to be taken to the hospital. You are required to take as many as possible, but you are not able to take everyone. Special obligations can here tell you to prioritize those with whom you have special bonds. You are required to take your partner, just as you are required to take your own victim to the hospital. This does adversely affect those who are not saved. However, even in the absence of special obligations they would not have been wronged had other people been selected.

Things are different, however, when other individuals have valid claims against you. Your special obligation to your partner would not license you to change the order of the waiting list for transplant kidneys that you administer. Others have a valid claim to you that you follow the procedure and allocate kidneys by reference to the waiting list. The question "why should it be fine for you to harm me because of *your* special ties" becomes salient. In the case of killing, it appears that it would be impermissible to kill an innocent bystander in order to save either oneself, or a close associate such as a partner or child, or even several of your children.³⁶ The best reason for this is that special obligations cannot override the valid claims of third parties. This reason applies to special obligations towards one's loved ones just as well as it applies to special obligations towards those one has previously wronged. If so, then the difference between letting be killed and letting be killed by oneself is morally

³⁶ E.g. Judith Jarvis Thomson, "Self-Defense," *Philosophy & Public Affairs* 20 (1991): 283-310, at pp. 289-91; and Thomas Hurka, "Proportionality and the Morality of War," *Philosophy & Public Affairs* 33 (2005): 34-66, at p. 60.

significant only in some contexts. Crucially, it is not significant in cases like the *Guilty Agent* when it comes to licensing harm.

Why is it that special obligations are limited in this way and cannot override the claims of others? The reason is that special obligations derive their force from our general obligation not to harm individuals. Special obligations arise in particular contexts where roles or conventions specify the expectations of individuals. Frustrating or disappointing these expectations would breach the general injunction against harming.³⁷ The argument that special obligations are local versions of our general obligation not to harm is most developed in the case of promising.³⁸ In the case of promising, several philosophers have argued that our requirement to keep our promises is based on the general requirement not to raise expectations that one later frustrates. The fact that special obligations are subsidiary to our general obligations explains both why immoral relationships, conventions or agreements cannot give rise to obligations, and why special obligations cannot override the valid claims of third parties.

Besides general considerations about special obligations, there are also reasons specific to the special obligation towards our own victims that indicate why this obligation should not license us to kill. The reason is that it is precisely the transgression of the constraint against killing that gave rise to the special obligation in the first place. There is something paradoxical about the idea that we can make up for our wrong by repeating what we just did. If someone asked us: “Why are you shooting this person?” our response “because I already shot some other people (and could use blood transfusions from my new victim to help my old victims)” seems odd. We would point out that the realization that the earlier act was wrong should give us a strong reason *not* to engage in the further act of shooting even more people. The obligation towards our victims is born out of a recognition and realization of the fact that we ought not harm. The obligation is subsidiary to the overall demand on us not to act in ways that impose serious harm.

³⁷ See Dworkin, *Justice for Hedgehogs*, pp. 300-17.

³⁸ E.g. Thomson, *The Realm of Rights*, pp. 294-321; Scanlon, *What We Owe to Each Other*, ch. 7; and Dworkin, *Justice for Hedgehogs*, pp. 303-11.

There is something incoherent about a morality which allows us to kill in order to save from killing. If it were permissible to save one person by killing someone else, then we would be allowed, for example, to point our gun towards the second person. But then by pointing our gun we have placed this person under as much danger as the first person was under. She would now have a claim that we ought to come to her rescue by killing someone else. This would mean our actions *replicate* the exact same situation we started with. Morally speaking we are in no way better off.³⁹

Now it may be objected that my argument works well only because I have considered easy cases so far in which we would be saving one person by killing another one. But in the cases involving deontological constraints the interesting feature is that we would be saving many people by killing someone else. What I want to suggest is that this feature is not relevant for our special obligation to make good for our wrong. Our obligation is directed to someone, it is not an impersonal value judgment but an obligation *owed to* someone. By killing the one you would honor your obligation towards each of the five separately. The special feature of your relation with your victims is one that holds only between your victim and you. This gives you a reason to treat the case as one of pairwise comparisons between each of your to-be victims and the one person you are about to kill. This pairwise comparison assess what you would be doing and the reasons you have for each option rather than comparing the harm the victims will incur. In this case this means a comparison between killing and allowing the death of one's own victim. If this is correct, then the aforementioned reasons hold. The fact that you can replicate the same reason five times does not matter since your obligation is owed to the individuals and not to the group. Therefore, your action has to be based on the various individual's complaints and not on complaints of groups.

My reasoning can also explain why we are not allowed to harm a new victim in ways that fall short of killing. For example, it seems plausible to me to say that we

³⁹ Quinn gives a similar example where the agent would be killing two in order to save one. Quinn, "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing," pp. 307-8.

should rather let our own victim die instead of mutilating someone else. I have argued earlier that the distinction between doing and allowing is strong enough to warrant a restriction on our actions when we could let more harm happen. This would be an example of such a case. I further denied that our special obligation towards our own victim makes a difference in cases of this kind.

My reasoning concerning special obligations also explains an asymmetry between cases of constraints where persons are involved and parallel cases of material goods. It would be wrong to kill the one to prevent ourselves from becoming the killer of the five. But it seems permissible to destroy one piece of art if this is the only way to prevent ourselves from destroying five pieces of art. (Imagine that you have set a bomb that is going to destroy five Margritte paintings. The only way to defuse the bomb is to shield them with one equally great painting by Monet.⁴⁰) In such a case we are confronted with impersonal value judgments and no longer with obligations owed to someone in particular.

To sum up, I agree that there is a difference between letting die (or be killed) and letting oneself become the killer. The difference is context sensitive. It can explain why we should prioritize aiding our own victims where we can. It can also explain why even on an agent-based justification for constraints we can be allowed to minimize our own wrongdoings in cases where material goods and not persons are involved. But I deny that the difference is significant enough to allow the killing of one in order to save five of one's own victims. Previously I justified the distinction between killing and letting die by appealing to the relation of the potential victims and the agent. The one person has a complaint not to be killed. The five have a complaint to be saved. They can add a complaint based on your special historic responsibility. They can complain that refraining from saving means refraining from saving them from your killing them. But I denied that this complaint would be strong enough.

⁴⁰ See also Kamm, *Morality, Mortality*, 2:241-42.

IV. Why Evaluate Actions One at a Time?

Thus far I have argued that the Guilty Agent is facing a problem of doing versus allowing in her second decision when deciding whether to place the sixth man over the bomb. If this is the case, then the agent-based rationale can be defended against the *Guilty Agent* case. There is one more caveat to the argument. In this statement I have tacitly assumed that we are justified in regarding the second action separately from the first one. But why should we regard them separately and not in conjunction with one another? It still remains true that with my second action, I can prevent myself from killing five people. I cannot undo my first action, but I can alter the results of my first action. So there is a sense in which whatever I will do, I will have killed either the one or the five. My argument has only shown that I will not have killed the five *with my second action*. Why should this matter? Why cannot the five complain to me that if I do not save them that I *will have killed* them?

We have a contrast here between two different questions that we can ask: (1) What should I do? (2) What should it be the case that I will have done? The two questions seem identical and they will in most cases yield the same answer. But in the cases I am considering they come apart since my second action can alter what can truthfully be said about my first action. Question (1) phrases the choice as between killing one or letting five people be killed by oneself. This is the question I have been answering so far. Question (2) phrases the choice as between having killed one person or having killed five persons. It asks us to evaluate our actions retrospectively. We should adopt a later point in time and evaluate what we will have done.⁴¹

⁴¹ The choice here resembles a choice that both Hanna and Ingmar Persson have argued is a challenge for how proponents of the Doctrine of Doing and Allowing should treat cases of allowing one's own doing. (Ingmar Persson, *From Morality to the End of Reason* (Oxford: Oxford University Press, 2013), pp. 102-5; Jason Hanna, "Enabling Harm, Doing Harm, and Undoing One's Own Behavior," *Ethics* 126 (2015): 68-90, at pp. 86-89; and Hanna, "Doing, Allowing, and the Moral Relevance of the Past," pp. 683-85.) Either we take a view that only the present action matters (akin to Question 1), in which case we cannot tell allowing one's own doing apart from standard cases of allowing. Alternatively, we focus on the reasons the agent has to ensure no one will be harmed (akin to Question 2), in which case we cannot account for intuitions about deontological constraints. The argument in my previous section, however, has shown that the first option can take into consideration the past. One's past actions can

I think the first question is the appropriate one. The most fundamental question of morality is “what should I do” or perhaps “what ought I to do”. Morality is a practical inquiry in our actions. It is a guide for action and asks how we should face and decide decision problems. It asks us to choose among the options currently available to us. And the options available to us are the different choices we can make. In the case of the *Guilty Agent* you choose how to act with our second action. Only derivatively it becomes true how you will have acted with our first action. You are not performing this first action, but only influencing its description and meaning. The relevant choice is between what you are doing with your new, second action.

While this view is seldom articulated, it has been recognized before. Judith Jarvis Thomson argues that in a case similar to the *Guilty Agent* case the perspective of the present action is relevant because the agent has to act in the present. It is the options that are available to the agent when deciding to act which count.⁴² T.M. Scanlon argues that it is a feature of moral principles that they can be employed as guides to deliberation. As such, the principles seek to answer the question “May one do X?”. Scanlon concedes that the question of permissibility can also be employed retrospectively or hypothetically. But the question of permissibility must also be *possibly* the object of a decision.⁴³ This indicates a close link between permissibility and the perspective of making a decision. Among other things, it means that the question of permissibility applies only to options that are in the choice set of an agent.⁴⁴

We can also see the appropriateness of the question “what should I do” by considering whether “ought” should be understood objectively. If ought should be understood objectively, then actual results of actions determine the permissibility of actions. An action which leads to harm through an unforeseeable and unpredictable process would be impermissible according to the objective ought. But this does not seem correct. Fluke consequences should not render otherwise innocuous actions

determine which special obligations one has now. If we can show that the first question is the right one to ask, then the challenge disappears.

⁴² Thomson, “The Trolley Problem,” pp. 1414-15.

⁴³ Scanlon, *Moral Dimensions*, pp. 9-10, 21-24.

⁴⁴ Scanlon, *Moral Dimensions*, pp. 58-59.

impermissible. This reason indicates that moral permissibility should be connected to a deliberative and action-guiding function.⁴⁵ In other words, it brings the question “what should I do” back in the focus.

On the other hand, the question “what should it be the case that I will have done” is misguided in thinking about which action to perform. Someone who approaches the decision in the *Guilty Agent* case thinking about whether or not she will become the killer of only one or of five people, is clearly asking the wrong question. Her question shows excessive self-concern for her own perspective. This is one way how an agent can show excessive self-concern in the face of moral decisions. The agent reasons that if she does not save the five, she will become a mass killer. And this, presumably, is worse than becoming a killer of one. Reasoning in such a way in order to decide what to do is insensitive to the moral problem at hand. Morality is about what we should do and what reasons we have for choosing among our actions. It is not about keeping one’s hands clean. Asking ourselves of which things it will be true that we have committed them appears to me like keeping a scorecard of one’s own moral record or collecting points to get into the good place. It does not show the right attitude of engagement with the moral dilemma we are facing at the moment. While agents can rightly reflect on their moral records, this reflection should not become decisive when thinking about actions that significantly impact others by possibly violating their rights.

This is not to deny that our past actions can matter, but they matter in a subtler way. We can acknowledge the importance of history without evaluating past and present together. What we did in the past can change our reasons for our future actions. I have already indicated one way how this may come about. We acquire a special obligation towards the people we have put in danger. But we can ask about the effect of this special obligation purely by reference to the new decision we now have to make. I have argued that in cases involving deontological constraints this special obligation would not be strong enough to change what we all-things-considered ought to do. But in other cases it will be. In deciding whether to help your

⁴⁵ Scanlon, *Moral Dimensions*, pp. 47-52 for a similar argument.

own victim or another victim, you are allowed to suspend impartiality between the two and attend to your own victim with priority.

I mentioned earlier that the second question asks for a kind of retrospective justification. It is important to distinguish the kind of retrospective justification that the question is asking for from more plausible candidates of retrospective justification. First, we might ask retrospectively what we should have done in the past. In this sense it is still a deliberative question. We reexamine the deliberation at the time of action, or perhaps we only now have time to deliberate whether our instinctive action was indeed right or wrong. Even though it is retrospective, we put ourselves in the position at the time of the decision. Second, future events may have an impact on the question whether or not we should feel regret or even be blameworthy for past wrongs. Bernard Williams gives the example of the painter Gauguin who abandons his family to go to paint in Tahiti. Gauguin's later success renders this abandonment the beginning of great artistic success. Perhaps, if this artistic success is also of sufficient moral value, Gauguin should not feel regret for this choice.⁴⁶ But neither of the two senses of retrospective justification is at play in the question "what should it be the case that I will have done". The question is not phrased as a mere restatement of the deliberative question, and the question is not about the appropriateness of reactive attitudes. Since neither of the two plausible readings of retrospective justification can be attached to question (2), we should regard it as the wrong question to ask.

This gives us now a good criterion for assessing the actions seriatim and in isolation. Note that this position responds to the worries that temporal restrictions on agent-relativity are not morally significant. I am convinced by the criticism that differences in time, at least small differences, do not carry moral significance. However, the criterion I have used is not a temporal one, but the criterion of being the same action. My stabbing in the *Inconclusive Agent* case will take several moments,

⁴⁶ B.A.O. Williams, "Moral Luck," *Proceedings of the Aristotelian Society* 50 (1976): 115-35, at pp. 117-23. See also Elizabeth Harman, "'I'll Be Glad I Did It' Reasoning and the Significance of Future Desires," *Philosophical Perspectives* 23 (2009): 177-99; R. Jay Wallace, *The View From Here* (Oxford: Oxford University Press, 2013), chs. 1-4; and Bernhard Salow, "Partiality and Retrospective Justification," *Philosophy & Public Affairs* 45 (2017): 8-26.

nevertheless it is the same action. The distinction between actions has sufficient moral significance to treat them separately as I have argued.

V. Self-Indulgence

I have already mentioned how thinking about “what should it be the case that I will have done” is unduly self-concerned. The focus of the question is on what it will be the case that can be said about the agent. Having killed five is then worse than having killed one. It is worse because being the killer of five is worse than being the killer of one. To ask this question, detached from the victims with whom the agent is in contact now, is to abstract away from them. It is to ask what happens to oneself as opposed to what one does to others. The question is a reminder of an attitude where the agent keeps a scorecard of her moral record or collects points to get into the good place.

While keeping a scorecard of one’s moral record is unduly self-concerned, it is not the only source of excessive self-concern. To think that we should not harm a person because it would mean that we have to get our hands dirty would be another example. As Nagel asks: “[What] gives one man a right to put the purity of his soul or the cleanness of his hands above the lives or welfare of large numbers of other people?”⁴⁷ This thought comes out most clearly in the kind of justification Kamm calls *agent-concerned*. But there is a concern that the objection generalizes to all agent-based justifications.⁴⁸ Because agent-based justifications focus on the agent as opposed to the victim, they raise a natural suspicion of being excessively self-concerned. The objection is that a justification why it is wrong to kill or violate rights ends up focusing exclusively on the killer and not on the person whose rights are at stake. Naturally, the question arises whether this objection applies to my construal of agent-based justifications as well.

⁴⁷ Nagel, *Mortal Questions*, p. 63.

⁴⁸ Scheffler, “Agent-Centered Restrictions, Rationality, and the Virtues,” pp. 415-17; and Kamm, *Morality, Mortality*, 2:249-52.

My agent-based justification started out with the special responsibility for our own actions. In this sense, the justification is linked to the agent. This special responsibility, I have argued, allows us to view the moral problem from a first-personal perspective. It asks you to assess the relation between your victim and you. The different relation that you have to the victim you are killing and to the victim that you are letting die explains the moral asymmetry between the two. The difference is not explained by the general fact that killing is (impersonally) worse than letting die. Nor is it explained by a reluctance of the agent to kill. We should draw a distinction between those agent-based justifications which rely on reasons related to the agent alone and those which rely on the relation between the agent and the victim. All three familiar categories of agent-based justifications (*agent-relative*, *agent-focused*, *agent-concerned*) are plausibly grouped together in the first group and contrast with my relational justification.

This relational understanding can help us respond to the charge of self-indulgence. The justification brings the victim into the picture. The constraint against killing the one exists not because of a feature of you, the agent, but rather because of your relation to your victim. This indicates that the justification is not unduly self-concerned but receptive to the fate of the victim. For example, my relational understanding would not rule out killing a person who wants to die. Justifications that emphasize that killing is impersonally bad or that agents should not get their hands dirty would seem to extend to these cases of voluntary euthanasia as well. These justifications are not receptive of the fate of the victim and can plausibly be charged with excessive self-concern. My justification, however, does not have this feature. Whether or not it is correct to charge non-relational agent-based justifications with self-indulgence is a question I cannot assess here. But what I have argued is that relational agent-based justifications like mine are not guilty of self-indulgence.

VI. Constraints and Non-Persons

A further criticism against agent-based justifications is that they are overbroad. They apply to the killing of persons and non-persons alike. Justifying why

persons and non-persons should be treated differently would indicate, Kamm argues, that it is actually a feature of the victim that makes the transgression of the constraint impermissible.⁴⁹ But there is also a motivation for not divorcing the case of persons and non-persons. In both cases, we can find examples where it seems that we should not do evil in order to bring good about (or prevent evil). There is some hesitation to pre-emptively bomb cultural heritage in war even if this bombing would demoralize the opponent who then will no longer destroy a slightly more valuable cultural heritage. Perhaps there is even a constraint against doing so. If this is so, a more unified explanation of these phenomena would be more satisfactory.⁵⁰

My justification can integrate these two concerns better than victim-based justifications can. I argued that we have a special responsibility for our own actions. This special responsibility allows us to view moral issues from a first-personal perspective. In the case of persons this means that we have to take into account the moral relation between us and our potential victims. But the special responsibility also holds when there are no persons involved. Given the absence of relational reasons, the difference between doing and allowing is less stark in cases involving non-persons. This explains the asymmetry between persons and non-persons. There will be fewer deontological constraints against wrongdoing that does not involve wronging. But it does so by appealing to a common justification for constraints in personal and non-personal cases.

The same difference appears again in cases where we can minimize our own wrong by performing the same action. The previous conduct has given rise to a special responsibility for our past actions. In the case of persons this takes the form of a moral obligation that is owed to our victims. I have argued that this directedness can help us explain deontological constraints. In the case of non-persons, the special responsibility is not directed. Therefore, there are fewer constraints against minimizing our own wrongs in the case of non-persons.⁵¹ Again, we observe an

⁴⁹ Kamm, *Morality, Mortality*, 2:241-42 and *Intricate Ethics*, p. 28.

⁵⁰ See also Michael Otsuka, "Kamm on the Morality of Killing," *Ethics* 108 (1997): 197-207, at p. 205.

⁵¹ I say fewer because the following judgment is still possible: Doing A is worse than letting oneself do B, where A and B are damage done to a non-person and B is a slightly greater

asymmetry between persons and non-persons that is grounded in a common justification.

VII. The Next Constraint You Come Up Against

In this chapter, I have outlined an agent-based justification for side constraints based on the distinction between doing and allowing. Unlike the familiar agent-relative, agent-focused or agent-concerned justification, my justification is relational. It emphasizes the moral relation the agent has with her would-be victims. I have responded to the *Guilty Agent* and the problem of how to account for a constraint against minimizing one's own violations. The *Guilty Agent* rests on a mistaken analysis of the choices the agent is facing. The case involves two separate actions and we have good reason to distinguish the moral evaluation of actions separately. If we do so, we can justify constraints in cases where minimizing one's own violations is at stake, too. My justification can also respond to the criticism of self-indulgence and can treat cases of non-persons better than victim-based accounts.

Constraints tell us that individuals have rights and there are things no person may do to them. We are constrained, in each action, simply by the next constraint that we come up against. There is nothing special about the one victim as opposed to the other victims. You simply encounter her right in the given situation. Kamm imagines a potential victim saying that "[it] is impermissible to treat people in certain ways and so it is not permissible to treat me in this way; I am simply the first person with this status that you came up against."⁵² I think this is a correct and very helpful way of understanding deontological constraints. Kamm justifies the constraining right by appealing to inviolability and high moral status. My justification shows that we do not need to make any such appeal if we want to justify that we are constrained by the next constraint we come up against. The one person has a right not to be killed. My argument that killing the one is harder to justify than letting five die establishes that

damage than A. I do not know whether there are any such cases, but nothing what I have said rules them out.

⁵² Kamm, *Morality, Mortality*, 2:248.

there is no exception to this right for the sake of saving the five. My further arguments that the distinctions between letting die and letting be killed, and between letting be killed and letting be killed by oneself, are not relevant in this case support this further. The arguments add that there cannot be any exception for the sake of minimizing violations of rights. It is the fact that we come up against the next constraining right that is the reason why we may not kill, but my argument does not need to rely on the idea of inviolability to make sense of this statement.

Bibliography

- Robert Merrihew Adams, "Should Ethics be More Impersonal? A Critical Note of Derek Parfit, *Reasons and Persons*," *Philosophical Review* 98 (1989): 438-84.
- Matthew D. Adler, "The Puzzle of "Ex Ante Efficiency": Does Rational Approvability Have Moral Weight?," *University of Pennsylvania Law Review* 151 (2003): 1255-90.
- , *Well-Being and Fair Distribution* (Oxford: Oxford University Press, 2011).
- G.E.M. Anscombe, *Intention*, 2nd edn. (Cambridge, MA.: Harvard University Press, 1963).
- , "Who Is Wronged? Philippa Foot on Double Effect: One Point, in *Elizabeth Anscombe, Human Life, Action and Ethics*, ed. Mary Geach and Luke Gormally (Exeter: Imprint Academic, 2005), pp. 249-51.
- Kenneth J. Arrow, *Social Choice and Individual Values*, 2nd edn. (New York: John Wiley & Sons, 1963).
- , "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice," *Journal of Philosophy* 70 (1973): 245-63.
- Elizabeth Ashford, "The Demandingness of Scanlon's Contractualism," *Ethics* 113 (2003): 273-302.
- Jamies Baillie, "Recent Work on Personal Identity," *Philosophical Books* 4 (1993): 193-206.
- Nuel Belnap and Mitchell Green, "Indeterminism and the Thin Red Line," *Philosophical Perspectives* 8 (1994): 365-88.
- Jonathan Bennett, *Events and their Names* (Oxford: Oxford University Press, 1988).
- Simon Blackburn, "Has Kant Refuted Parfit?," in *Reading Parfit*, ed. Jonathan Dancy (Oxford: Blackwell, 1997), pp. 180-201.
- Andrew Brennan, "Amnesia and Psychological Continuity," *Canadian Journal of Philosophy* 15 (1985): 195-209.
- David O. Brink, "Sidgwick and the Rationale for Rational Egoism," in *Essays on Henry Sidgwick*, ed. Bart Schultz (Cambridge: Cambridge University Press, 1992), pp. 199-240.
- , "The Separateness of Persons, Distributive Norms, and Moral Theory," in *Value, Welfare, and Morality*, ed. R.G. Frey and Christopher Morris (Cambridge: Cambridge University Press, 1993), pp. 252-89.
- , "Rational Egoism and the Separateness of Persons," in Dancy, *Reading Parfit*, pp. 96-134.
- , "Self-Love and Altruism," *Social Philosophy & Policy* 14 (1997): 122-57.
- Richard Brook, "Agency and Morality," *Journal of Philosophy* 88 (1991): 190-212.
- John Broome, "Fairness," *Proceedings of the Aristotelian Society* 91 (1991): 87-102.

- , "Utilitarian Metaphysics," in *Interpersonal Comparisons of Well-Being*, ed. Jon Elster and John E. Roemer (Cambridge: Cambridge University Press, 1991), pp. 70-97.
- Campbell Brown, "Is close enough good enough?," *Economics & Philosophy* 36 (2020): 29-59.
- Lara Buchak, *Risk and Rationality* (Oxford: Oxford University Press, 2013).
- , "Taking Risks behind the Veil of Ignorance," *Ethics* 127 (2017): 610-44.
- Susanne Burri, "Personal Sovereignty and Our Moral Rights to Non-Interference," *Journal of Applied Philosophy* 34 (2017): 621-34.
- Rudolf Carnap, "The Two Concepts of Probability," *Philosophy and Phenomenological Research* 5 (1945): 513-32.
- Richard Yetter Chappell, "Value Receptacles," *Noûs* 49 (2015): 322-32.
- Tim Christie, "Natural Separateness: Why Parfit's Reductionist Account of Persons Fails to Support Consequentialism," *Journal of Moral Philosophy* 6 (2009): 178-95.
- G.A. Cohen, *Rescuing Justice and Equality* (Cambridge, MA.: Harvard University Press, 2008).
- , "Rescuing Conservatism: A Defense of Existing Value," in *Finding Oneself in the Other*, ed. Michael Otsuka (Princeton and Oxford: Princeton University Press, 2013), pp. 143-74.
- Norman Daniels, "Can There be Moral Force to Favoring an Identified over a Statistical Life?," in *Identified versus Statistical Lives*, ed. I. Glenn Cohen, Norman Daniels, and Nir Eyal (Oxford: Oxford University Press, 2015), pp. 110-23.
- Stephen Darwall, "Agent-Centered Restrictions from the Inside Out," *Philosophical Studies* 50 (1986): 291-319.
- , *The Second-Person Standpoint* (Cambridge, MA.: Harvard University Press, 2006).
- Donald Davidson, "Actions, Reasons, and Causes," *Journal of Philosophy* 60 (1963): 685-700.
- , "The Individuation of Events," in *Essays in Honor of Carl G. Hempel*, ed. Nicholas Rescher (Dordrecht: Reidel, 1969), pp. 295-309.
- , *Essays on Actions and Events* (Oxford: Oxford University Press, 2001).
- Tyler Doggett, "Saving the Few," *Noûs* 47 (2013): 302-15.
- Dale Dorsey, "Headaches, Lives and Value," *Utilitas* 21 (2009): 36-58.
- Lawrence Davis, "Individuation of Action," *Journal of Philosophy* 67 (1970): 520-30.
- Tom Dougherty, "Aggregation, Beneficence and Chance," *Journal of Ethics and Social Philosophy* 7 (2013): 1-19.
- Julia Driver, "The History of Utilitarianism," in *The Stanford Encyclopedia of Philosophy*. Winter 2014 Edition, ed. Edward N. Zalta, URL: <<https://plato.stanford.edu/archives/win2014/entries/utilitarianism-history/>>.
- Ronald Dworkin, *Taking Rights Seriously* (London: Duckworth, 1978).

- , *Sovereign Virtue* (Cambridge, MA.: Harvard University Press, 2000).
- , *Justice for Hedgehogs* (Cambridge, MA.: The Belnap Press of Harvard University Press).
- Antony Eagle, "Deterministic Chance," *Noûs* 45 (2011): 269-99.
- , "Chance versus Randomness," in *The Stanford Encyclopedia of Philosophy*. Spring 2019 Edition, ed. Edward N. Zalta, URL: <<https://plato.stanford.edu/archives/spr2019/entries/chance-randomness/>>.
- Nina Emery, "Chance, Possibility, and Explanation," *British Journal for the Philosophy of Science* 66 (2015): 95-120.
- Nir Eyal, "Concentrated Risk, the Coventry Blitz, Chamberlain's Cancer," in Cohen, Daniels, and Eyal, *Identified versus Statistical Lives*, pp. 94-109.
- Joel Feinberg, *Rights, Justice, and the Bounds of Liberty* (Princeton: Princeton University Press, 1980).
- John Findlay, *Values and Intentions* (London: George Allen & Unwin, 1961).
- William J. FitzPatrick, "Acts, intentions, and moral permissibility: in defence of the doctrine of double effect," *Analysis* 63 (2003): 317-21.
- Marc Fleurbaey, "Equality of Resources Revisited," *Ethics* 113 (2002): 82-105.
- , "Economics and Economic Justice," in *The Stanford Encyclopedia of Philosophy*. Winter 2016 Edition, ed. Edward N. Zalta, URL: <<https://plato.stanford.edu/archives/win2016/entries/economic-justice/>>.
- Marc Fleurbaey and Alex Voorhoeve, "Decide As You Would with Full Information!," in *Inequalities in Health*, ed. Nir Eyal, Samia A. Hurst, Ole F. Norheim, and Dan Wikler (Oxford: Oxford University Press, 2013), pp. 113-28.
- Harry Frankfurt, "The Importance of What We Care About," *Synthese* 53 (1982): 257-72.
- Samuel Freeman, *Justice and the Social Contract* (Oxford: Oxford University Press, 2007).
- Johann Frick, "Uncertainty and Justifiability to Each Person. Response to Fleurbaey and Voorhoeve," in Eyal, Hurst, Norheim, and Wikler, *Inequalities in Health*, pp. 129-46.
- , "Contractualism and Social Risk," *Philosophy & Public Affairs* 43 (2015): 175-223.
- Barbara H. Fried, "Can Contractualism Save Us from Aggregation?," *The Journal of Ethics* 16 (2012): 39-66.
- Roman Frigg and Carl Hoefer, "The Best Humean System for Statistical Mechanics," *Erkenntnis* 80 (2015): 551-74.
- David Gauthier, *Practical Reasoning* (Oxford: Clarendon Press, 1963).
- , *Morals by Agreement* (Oxford: Oxford University Press, 1986).
- Jonathan Glover, *Causing Death and Saving Lives* (London: Penguin Books, 1977).

Luke Glynn, "Deterministic Chance," *British Journal for the Philosophy of Science* 61 (2010): 51-80.

Peter Godfrey-Smith, *Other Minds* (London: William Collins, 2017).

---, "Evolving Across the Explanatory Gap," *Philosophy, Theory, and Practice in Biology* 11 (2019): 1-24.

Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs: Prentice-Hall Inc., 1970).

Hilary Greaves, "A Reconsideration of the Harsanyi-Sen-Weymark Debate on Utilitarianism," *Utilitas* 29 (2017): 175-213.

Hilary Greaves and Harvey Lederman, "Extended Preferences and Interpersonal Comparisons of Well-Being," *Philosophy and Phenomenological Research* 96 (2018): 636-67.

Geoffrey Russell Grice, *The Grounds of Moral Judgement* (Cambridge: Cambridge University Press), 1967.

Alan Hájek, "Interpretations of Probability," in *The Stanford Encyclopedia of Philosophy*. Fall 2019 Edition, ed. Edward N. Zalta, URL: <<https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>>.

Vinit Haksar, *Equality, Liberty, and Perfectionism* (Oxford: Oxford University Press, 1979).

---, *Indivisible selves and moral practice* (Edinburgh: Edinburgh University Press, 1991).

John Halstead, "The Numbers Always Count," *Ethics* 126 (2016): 789-802.

Jason Hanna, "Enabling Harm, Doing Harm, and Undoing One's Own Behavior," *Ethics* 126 (2015): 68-90.

---, "Doing, Allowing, and the Moral Relevance of the Past," *Journal of Moral Philosophy* 12 (2015): 677-98.

Caspar Hare, "Obligation and Regret When There is No Fact of the Matter About What Would Have Happened if You Had not Done What You Did," *Noûs* 45 (2011): 190-206.

---, "Obligations to Merely Statistical People," *The Journal of Philosophy* 109 (2012): 378-90.

---, "Should We Wish Well to All?," *Philosophical Review* 125 (2016): 451-72.

R.M. Hare, *Freedom and Reason* (Oxford: Clarendon Press, 1963).

Elizabeth Harman, "'I'll Be Glad I Did It' Reasoning and the Significance of Future Desires," *Philosophical Perspectives* 23 (2009): 177-99.

John C. Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy* 61 (1953): 434-35.

---, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63 (1955): 309-21.

- , "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory," *American Political Science Review* 69 (1975): 594-606.
- , *Rational behavior and bargaining equilibrium in games and social situations* (Cambridge: Cambridge University Press, 1977).
- , "Morality and the theory of rational behavior," in *Utilitarianism and beyond*, ed. Amartya Sen and Bernard Williams (Cambridge: Cambridge University Press, 1982), pp. 39-62.
- Ulrike Heuer, "The Paradox of Deontology, Revisited," in *Oxford Studies in Normative Ethics. Volume 1*, ed. Mark Timmons (Oxford: Oxford University Press, 2011), pp. 237-67.
- Iwao Hirose, "Saving the greater number without combining claims," *Analysis* 61 (2001): 341-43.
- , *Moral Aggregation* (Oxford: Oxford University Press, 2014).
- Carl Hoefer, "The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance," *Mind* 116 (2007): 549-96.
- Joe Horton, "Aggregation, Complaints, and Risk," *Philosophy & Public Affairs* 45 (2017): 54-81.
- , "Always Aggregate," *Philosophy & Public Affairs* 46 (2018): 160-74.
- David Hume, *A Treatise of Human Nature*, ed. L.A. Selby-Bigge (Oxford: Clarendon Press, 1888).
- Thomas Hurka, "Proportionality and the Morality of War," *Philosophy & Public Affairs* 33 (2005): 34-66.
- Paul Hurley, "Agent-Centered Restrictions: Clearing the Air of Paradox," *Ethics* 108 (1997): 120-46.
- Aaron James, "Contractualism's (Not So) Slippery Slope," *Legal Theory* 18 (2012): 263-92.
- S.D. John, "Risk, Contractualism, and Rose's 'Prevention Paradox'," *Social Theory and Practice* 40 (2014): 28-50.
- Christa M. Johnson, "The Intrapersonal Paradox of Deontology," *Journal of Moral Philosophy* 16 (2019): 279-301.
- Mark Johnston, "Reasons and Reductionism," *Philosophical Review* 101 (1992): 589-618.
- , "Human Concerns without Superlative Selves," in Dancy, *Reading Parfit*, pp. 149-79.
- Shelly Kagan, "Replies to my Critics," *Philosophy and Phenomenological Research* 51 (1991): 919-28.
- F.M. Kamm, "Equal Treatment and Equal Chances," *Philosophy & Public Affairs* 14 (1985): 177-94.
- , *Morality, Mortality*, vol. 1 (Oxford: Oxford University Press, 1993).

- , *Morality, Mortality*, vol. 2 (New York: Oxford University Press, 1996).
- , "Moral Status and Personal Identity: Clones, Embryos, and Future Generations," *Social Philosophy & Policy* 22 (2005): 283-307.
- , *Intricate Ethics* (Oxford: Oxford University Press, 2007).
- , *Bioethical Prescriptions* (Oxford: Oxford University Press, 2013).
- Serge-Christophe Kolm, *Justice and Equity* (Cambridge, MA.: The MIT Press, 1997).
- Christine M. Korsgaard, "Two Distinctions in Goodness," *Philosophical Review* 92 (1983): 169-195.
- , "Personal Identity and the Unity of Agency: A Kantian Response to Parfit," *Philosophy & Public Affairs* 18 (1989): 101-32.
- Rahul Kumar, "Contractualism on saving the many," *Analysis* 61 (2001): 165-70.
- , "Risking and Wronging," *Philosophy & Public Affairs* 43 (2015): 27-51.
- A.R. Lacey, "Sidgwick's Ethical Maxims," *Philosophy* 34 (1959): 217-28.
- Anthony Simon Laden, "Taking the Distinction between Persons Seriously," *Journal of Moral Philosophy* 1 (2004): 277-92.
- Seth Lazar, "Limited Aggregation and Risk," *Philosophy & Public Affairs* 46 (2018): 117-59.
- Seth Lazar and Chad Lee-Stronach, "Axiological Absolutism and Risk," *Noûs* 53 (2019): 97-113.
- David Lefkowitz, "On the Concept of a Morally Relevant Harm," *Utilitas* 20 (2008): 409-23.
- James Lenman, "Contractualism and risk imposition," *Politics, Philosophy & Economics* 7 (2008): 99-122.
- Clarence Irving Lewis, *An Analysis of Knowledge and Valuation* (La Salle: The Open Court Publishing Company, 1946).
- C.S. Lewis, *The Problem of Pain* (Québec: Samizdat University Press, 2016).
- David Lewis, "Survival and Identity," in *The Identities of Persons*, ed. Amélie Oksenberg Rorty (Berkeley: University of California Press, 1976), pp. 17-40.
- , "A Subjectivist's Guide to Objective Chance," in *Studies in Inductive Logic and Probability*, vol. 2, ed. Richard C. Jeffrey (Berkeley: University of California Press, 1980), pp. 263-93.
- , *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986).
- S. Matthew Liao, "Who Is Afraid of Numbers?," *Utilitas* 20 (2008): 447-61.
- Kasper Lippert-Rasmussen, "Moral Status and the Impermissibility of Minimizing Violations," *Philosophy & Public Affairs* 25 (1996): 333-51.
- Christian List and Marcus Pivato, "Emergent Chance," *Philosophical Review* 124 (2015): 119-52.

- Barry Loewer, "Determinism and Chance," *Studies in History and Philosophy of Modern Physics* 32 (2001): 609-20.
- Weyma Lübbe, "Taurek's No Worse Claim," *Philosophy & Public Affairs* 36 (2008): 68-85.
- Aidan Lyon, "Deterministic probability: neither chance nor credence," *Synthese* 182 (2011): 413-32.
- James MacKaye, *The Economy of Happiness* (Boston: Little, Brown, and Company, 1906).
- Anna Mahtani, "The Ex Ante Pareto Principle," *The Journal of Philosophy* 114 (2017): 303-23.
- C.B. Martin and Max Deutscher, "Remembering," *Philosophical Review* 75 (1966): 161-96.
- Dennis McKerlie, "Egalitarianism and the Difference Between Interpersonal and Intrapersonal Judgments," in *Egalitarianism*, ed. Nils Holtug and Kasper Lippert-Rasmussen (Oxford: Clarendon Press, 2006), pp. 157-73.
- Jeff McMahan, *The Ethics of Killing* (Oxford: Oxford University Press, 2002).
- Christopher McMahon, "The Paradox of Deontology," *Philosophy & Public Affairs* 20 (1991): 350-77.
- David McNaughton and Piers Rawling, "On Defending Deontology," *Ratio* 11 (1998): 37-54.
- Alfred R. Mele, *Springs of Action* (Oxford: Oxford University Press, 1992).
- Michael Moehler, "The Rawls-Harsanyi Dispute: A Moral Point of View," *Pacific Philosophical Quarterly* 99 (2018): 82-99.
- Andreas L. Mogensen, "The Brave Officer Rides Again," *Erkenntnis* 83 (2018): 315-29.
- Philippe Mongin, "The Impartial Observer Theorem of Social Ethics," *Economics & Philosophy* 17 (2001): 147-79.
- Véronique Munoz-Dardé, "The Distribution of Numbers and the Comprehensiveness of Reasons," *Proceedings of the Aristotelian Society*, 105 (2005): 191-217.
- Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970).
- , *Mortal Questions* (Cambridge: Cambridge University Press, 1979).
- , *The View From Nowhere* (New York: Oxford University Press, 1986).
- , *Equality and Partiality* (Oxford: Oxford University Press, 1991).
- , "Personal Rights and Public Space," *Philosophy & Public Affairs* 24 (1995): 83-107.
- Alastair Norcross, "Comparing Harms: Headaches and Human Lives," *Philosophy & Public Affairs* 26 (1997): 135-67.

- , "Two Dogmas of Deontology: Aggregation, Rights, and the Separateness of Persons," *Social Philosophy & Policy* 26 (2009): 76-95.
- Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974).
- Martha C. Nussbaum, *Sex and Social Justice* (New York: Oxford University Press, 1999).
- John Oberdiek, *Imposing Risk* (Oxford: Oxford University Press, 2017).
- Eric Olson, "Why Definitions of Death Don't Matter," (unpublished manuscript).
- Michael Otsuka, "Kamm on the Morality of Killing," *Ethics* 108 (1997): 197-207.
- , "Saving Lives, Moral Theory, and the Claims of Individuals," *Philosophy & Public Affairs* 34 (2006): 109-35.
- , "Are deontological constraints irrational?," in *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, ed. Ralf M. Bader and John Meadowcroft (Cambridge: Cambridge University Press, 2011), pp. 38-58.
- , "Risking Life and Limb: How to Discount Harms by Their Improbability," in Cohen, Daniels, Eyal, *Identified versus Statistical Lives*, pp. 77-93.
- , "Personal Identity, Substantial Change, and the Significance of Becoming," *Erkenntnis* 83 (2018): 1229-43.
- , "Determinism and the Value and Fairness of Equal Chances," (unpublished manuscript).
- 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 (2009): 171-99.
- Derek Parfit, "Personal Identity," *Philosophical Review* 80 (1971): 3-27.
- , "On 'The Importance of Self-Identity'," *Journal of Philosophy* 68 (1971): 683-90.
- , "Later selves and moral principles," in *Philosophy and Personal Relations*, ed. Alan Montefiore (London: Routledge, 1973), pp. 137-69.
- , "Lewis, Parry, and What Matters," in Rorty, *The Identities of Persons*, pp. 91-108.
- , "Innumerate Ethics," *Philosophy & Public Affairs* 7 (1978): 285-301.
- , *Reasons and Persons* (Oxford: Clarendon Press, 1984).
- , "The Unimportance of Identity," in *Identity*, ed. Henry Harris (Oxford: Oxford University Press, 1995), pp. 13-45.
- , "Experiences, Subjects, and Conceptual Schemes," *Philosophical Topics* 26 (1999): 217-70.
- , "Justifiability to Each Person," *Ratio* 16 (2003): 368-90.
- , *On What Matters*, vol. 1 (Oxford: Oxford University Press, 2011).
- , *On What Matters*, vol. 2 (Oxford: Oxford University Press, 2011).
- L.A. Paul, *Transformative Experiences* (Oxford: Oxford University Press, 2014).

- John Perry, "The Importance of Being Identical," in Rorty, *The Identities of Persons*, pp. 67-90.
- Ralph Barton Perry, *General Theory of Value* (Cambridge, MA.: Harvard University Press, 1926).
- Ingmar Persson, *From Morality to the End of Reason* (Oxford: Oxford University Press, 2013).
- Richard Price, *A Review of the Principal Questions in Morals* (Oxford: Clarendon Press, 1948).
- Warren S. Quinn, "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing," *Philosophical Review* 98 (1989): 287-312.
- Anatol Rapoport, *Strategy and Conscience* (New York: Harper & Row, 1964).
- John Rawls, "Justice as Fairness," *Philosophical Review* 67 (1958): 164-94.
- , *A Theory of Justice* (Cambridge, MA.: Harvard University Press, 1971).
- , "Some Reasons for the Maximin Criterion," *American Economic Review* 64 (1974): 141-46.
- , "Social unity and primary goods," in Sen and Williams, *Utilitarianism and beyond*, pp. 159-85.
- , "Justice as Fairness: Political not Metaphysical," *Philosophy & Public Affairs* 14 (1985): 223-51.
- , "Constitutional Liberty and the Concept of Justice," in *John Rawls: Collected Papers*, ed. Samuel Freeman (Cambridge, MA.: Harvard University Press, 1999), pp. 73-95.
- , *A Theory of Justice*, rev. edn. (Oxford: Oxford University Press, 1999).
- , *Lectures on the History of Moral Philosophy* (Cambridge, MA.: Harvard University Press, 2000).
- , *Justice as Fairness. A Restatement* (Cambridge, MA.: Harvard University Press, 2001).
- Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986).
- Tom Regan, *The Case for Animal Rights* (Berkeley: University of California Press, 1983).
- , "The Case for Animal Rights," in *In Defense of Animals*, ed. Peter Singer (New York: Basil Blackwell, 1985), pp. 13-26.
- Sophia Reibetanz, "Contractualism and Aggregation," *Ethics* 108 (1998): 296-311.
- John E. Roemer, "Equality of Talent," *Economics & Philosophy* 1 (1985): 151-88.
- , "Harsanyi's Impartial Observer is not a Utilitarian," in *Justice, Political Liberalism, and Utilitarianism*, ed. Marc Fleurbaey, Maurice Salles, and John Weymark (Cambridge: Cambridge University Press, 2008), pp. 129-35.
- Thomas Rowe, "Risk and the Unfairness of Some Being Better Off at the Expense of Others," *Journal of Ethics and Social Philosophy* 16 (2019): 44-66.

- Korbinian Rüger, "On Ex Ante Contractualism," *Journal of Ethics and Social Philosophy* 13 (2018): 240-258.
- , "Aggregation with Constraints," *Utilitas* (forthcoming).
- Bernhard Salow, "Partiality and Retrospective Justification," *Philosophy & Public Affairs* 45 (2017): 8-26.
- Michael Sandel, *Liberalism and the Limits of Justice*, 2nd edn. (Cambridge: Cambridge University Press, 1998).
- T.M. Scanlon, "Preference and Urgency," *Journal of Philosophy* 72 (1975): 655-69.
- , "Contractualism and utilitarianism," in Sen and Williams, *Utilitarianism and beyond*, pp. 103-28.
- , *What We Owe to Each Other* (Cambridge, MA.: The Belknap Press of Harvard University Press, 1998).
- , *Moral Dimensions* (Cambridge, MA.: The Belknap Press of Harvard University Press, 2008).
- , "Reply to Zofia Stemplowska," *Journal of Moral Philosophy* 10 (2013): 508-14.
- , "Contractualism and Justification," (unpublished manuscript).
- Marya Schechtman, *The Constitution of Selves* (Ithaca: Cornell University Press, 1996).
- Samuel Scheffler, *The Rejection of Consequentialism* (Oxford: Oxford University Press, 1982).
- , "Agent-Centered Restrictions, Rationality, and the Virtues," *Mind* 94 (1985): 409-19.
- , "Rawls and Utilitarianism," in *The Cambridge Companion to Rawls*, ed. Samuel Freeman (Cambridge: Cambridge University Press, 2003), pp. 426-59.
- John R. Searle, *Intentionality* (Cambridge: Cambridge University Press, 1983).
- Shlomi Segall, "Sufficientarianism and the Separateness of Persons," *Philosophical Quarterly* 69 (2019): 142-55.
- Amartya Sen, "Welfare Inequalities and Rawlsian Axiomatics," *Theory and Decision* 7 (1976): 243-62.
- , "Non-Linear Social Welfare Functions: A Reply to Professor Harsanyi," in *Foundational Problems in the Special Sciences*, ed. Robert E. Butts and Jaakko Hintikka (Dordrecht: D. Reidel, 1977), pp. 297-302.
- , "Utilitarianism and Welfarism," *Journal of Philosophy* 76 (1979): 463-89.
- George Sher, "What Makes a Lottery Fair?," *Noûs* 14 (1980): 203-16.
- David W. Shoemaker, "Theoretical Persons and Practical Agents," *Philosophy & Public Affairs* 25 (1996): 318-32.
- , "Selves and Moral Units," *Pacific Philosophical Quarterly* 80 (1999): 391-419.

---, "Parfit's 'Argument from Below' vs. Johnston's 'Argument from Above'," in *PEA Soup Blog* (2006, URL: <https://peasoup.typepad.com/peasoup/2006/04/parfits_argumen.html>).

---, "Personal Identity and Ethics," in *The Stanford Encyclopedia of Philosophy*. Winter 2016 Edition, ed. Edward N. Zalta (URL: <<https://plato.stanford.edu/archives/win2016/entries/identity-ethics/>>).

Sydney Shoemaker, "Persons and Their Pasts," *American Philosophical Quarterly* 7 (1970): 269-85.

---, "Personal Identity: A materialist's account," in *Personal Identity*, ed. Sydney Shoemaker and Richard Swinburne (Oxford: Basil Blackwell, 1984), pp. 67-132.

Alan Sidelle, "Parfit on 'the Normal/a Reliable/any Cause of Relation R,'" *Mind* 120 (2011): 735-60.

Henry Sidgwick, *The Methods of Ethics* (Chicago: University of Chicago Press, 1962).

Thomas Sinclair, "Are We Conditionally Obligated to be Effective Altruists?," *Philosophy & Public Affairs* 46 (2018): 36-59.

J.J.C. Smart, *An Outline of a System of Utilitarian Ethics* (London and New York: Cambridge University Press, 1961).

Adam Smith, "The Theory of Moral Sentiments," in *The Essential Adam Smith*, ed. Robert L. Heilbroner (Oxford: Oxford University Press, 1986), pp. 57-147.

Elliott Sober, "Evolutionary Theory and the Reality of Macro-Probabilities," in *The Place of Probability in Science*, ed. Ellery Eells and J.H. Fetzer (Dordrecht: Springer, 2010), pp. 133-61.

Ernest Sosa, "Surviving Matters," *Noûs* 24 (1990): 297-322.

Kai Spiekermann, "Good Reasons for Losers: Lottery Fairness and Social Risk," (unpublished manuscript).

H. Orri Stefánsson and Richard Bradley, "What is Risk Aversion?," *British Journal for the Philosophy of Science* 70 (2019): 77-102.

Michael Strevens, "Probability out of Determinism," in *Probabilities in Physics*, ed. Claus Beisbart and Stephan Hartmann (Oxford: Oxford University Press, 2011), 339-64.

Victor Tadros, "Controlling Risk," in *Prevention and the Limits of the Criminal Law*, ed. Andrew Ashworth, Lucia Zedner, and Patrick Tomlin (Oxford: Oxford University Press, 2013), pp. 133-55.

---, "Localized Restricted Aggregation," in *Oxford Studies in Political Philosophy. Volume 5*, ed. David Sobel, Peter Vallentyne, and Steven Wall (Oxford: Oxford University Press, 2019), pp. 171-204.

John M. Taurek, "Should the Numbers Count?," *Philosophy & Public Affairs* 6 (1977): 293-316.

Larry Temkin, *Rethinking the Good* (Oxford: Oxford University Press, 2012).

- Judith Jarvis Thomson, "The Time of a Killing," *Journal of Philosophy* 68 (1971): 115-32.
- , "The Trolley Problem," *Yale Law Journal* 94 (1985): 1395-1415.
- , *The Realm of Rights* (Cambridge, MA.: Harvard University Press, 1990).
- , "Self-Defense," *Philosophy & Public Affairs* 20 (1991): 283-310.
- , "Turning the Trolley," *Philosophy & Public Affairs* 36 (2008): 359-74.
- Jan Tinbergen, "Welfare Economics and Income Distribution," *American Economic Review* 47 (1957): 490-503.
- Patrick Tomlin, "On Limited Aggregation," *Philosophy & Public Affairs* 45 (2017): 232-60.
- Jason Tyndal, "The Separateness of Persons: A Moral Basis for a Public Justification Requirement," *Journal of Value Inquiry* 51 (2017): 491-505.
- Peter Unger, *Identity, Consciousness and Value* (New York: Oxford University Press, 1990).
- Aart van Gils and Patrick Tomlin, "Relevance Rides Again?," in *Oxford Studies in Political Philosophy. Volume 6*, ed. David Sobel, Peter Vallentyne, and Steven Wall (Oxford: Oxford University Press, 2020), pp. 221-55.
- J. David Velleman, "Self to Self," *Philosophical Review* 105 (1996): 39-76.
- William Vickrey, "Measuring Marginal Utility by Reactions to Risk," *Econometrica* 13 (1945): 319-33.
- Alex Voorhoeve, "How Should We Aggregate Competing Claims?," *Ethics* 125 (2014): 64-87.
- , "Matthew D. Adler: Well-being and fair distribution: beyond cost-benefit analysis," *Social Choice and Welfare* 42 (2014): 245-54.
- , "Why One Should Count Only Claims with which One Can Sympathize," *Public Health Ethics* 10 (2017): 148-56.
- , "Balancing small against large burdens," *Behavioural Public Policy* 2 (2018): 125-42.
- , "May a Government Mandate more Comprehensive Health Insurance than Citizens Want for Themselves?," in *Oxford Studies in Political Philosophy. Volume 4*, ed. David Sobel, Peter Vallentyne, and Steven Wall (Oxford: Oxford University Press, 2018), pp. 167-91.
- R. Jay Wallace, *The View From Here* (Oxford: Oxford University Press, 2013).
- , *The Moral Nexus* (Princeton: Princeton University Press, 2019).
- Alec Walen, "Doing, Allowing, and Disabling: Some Principles Governing Deontological Restrictions," *Philosophical Studies* 80 (1995): 183-215.
- , "Risks and Weak Aggregation: Why Different Models of Risk Suit Different Types of Cases," *Ethics* (forthcoming).

David Wasserman and Alan Strudler, "Can a Nonconsequentialist Count Lives?," *Philosophy & Public Affairs* 31 (2003): 71-94.

J.W.N. Watkins, "Towards a Unified Decision Theory: A Non-Bayesian Approach," in Butts and Hintikka, *Foundational Problems in the Special Sciences*, pp. 345-79.

John A. Weymark, "A reconsideration of the Harsanyi-Sen debate on utilitarianism," in Elster, Roemer, *Interpersonal Comparisons of Well-Being*, pp. 255-320.

Jennifer Whiting, "Friends and Future Selves," *Philosophical Review* 95 (1986): 547-80.

Bernard Williams, "A Critique of Utilitarianism," in *Utilitarianism: For and Against*, ed. J.J.C. Smart and Bernard Williams (Oxford: Oxford University Press, 1973), pp. 77-150.

---, *Problems of the Self* (Cambridge: Cambridge University Press, 1973).

---, "Moral Luck," *Proceedings of the Aristotelian Society* 50 (1976): 115-35.

---, "Persons, Character and Morality," in Rorty, *The Identities of Persons*, pp. 197-216

Susan Wolf, "Self-Interest and Interest in Selves," *Ethics* 96 (1986): 704-20.

Alan Zaitchik, "Trammell on Positive and Negative Duties," *The Personalist* 58 (1977): 93-96.

Matt Zwolinski, "The Separateness of Persons and Liberal Theory," *Journal of Value Inquiry* 42 (2008): 147-65.