

**EQUAL OPPORTUNITY,
EQUALITY,
AND RESPONSIBILITY**

SUBMITTED BY

ALEXANDER EDMUND VOORHOEVE

OF

THE DEPARTMENT OF PHILOSOPHY
OF UNIVERSITY COLLEGE LONDON

FOR THE

PH.D. IN PHILOSOPHY

Abstract

This thesis argues that a particular version of equal opportunity for welfare is the best way of meeting the joint demands of three liberal egalitarian ideals: distributional equality, responsibility, and respect for individuals' differing reasonable judgements of their own good. It also examines which social choice rules best represent these demands. Finally, it defends the view that achieving equal opportunity for welfare should not only be a goal of formal public institutions, but that just citizens should also sometimes be guided by it in their everyday life.

The version of equal opportunity for welfare it defends differs from some well-known contemporary versions in the following ways. First, it rejects a definition of welfare as the degree of satisfaction of a person's preferences, because, it argues, this conception of welfare cannot adequately deal with preference change. Instead, it suggests that we should adopt a conception of welfare based on a list of goods and conditions that are recognised as valuable from the perspective of a variety of different conceptions of the good.

Second, it argues that individuals' *prima facie* claim to an equally valuable share of the world's resources—a claim which is based on their equal moral worth—is limited to situations in which giving one person a more valuable share means that someone else ends up with a less valuable share. It also argues that in situations where we can improve at least one person's situation without worsening anyone else's, we generally do not fail to respect each person's equal moral worth by doing so, even if this leads to inequalities.

Third, it defends a distinct view of responsibility, which justifies social arrangements that give people certain options with reference to the value that individuals can achieve (but don't necessarily achieve) through their choices from these options.

Acknowledgements

“What, besides binge drinking—so harmful to one’s mind and health—does Holland have to offer you? Go to Britain, that country made famous by the writings of its many scholars. Time spent with its great thinkers, its extraordinarily cultivated men, will civilise you, and make you worldly-wise.”¹

I received this advice from an elderly Erasmus to a Dutch student of his day upon my graduation from the University of Rotterdam, just before my move to Britain. A friend had given me a collection of Erasmus’s letters as a going away present. These letters contained little to encourage me to take up a scholarly career. Erasmus complained to his friends about dull students and boring colleagues, about the poverty that forced him to supplicate the Church for scholarships, and reported how philosophical disagreements turned colleagues into bitter enemies. I soon discovered, however, that the tone changed whenever he wrote about Britain. “If you knew what was good for you”, he wrote a Parisian friend, Fausto Andrelini, during his first trip to Britain in the Spring of 1499, “you’d come here as soon as possible. Why should a man of your calibre grow old in French muck? (...) If you knew the true nature of Britain’s enjoyments, you’d make your way here on winged feet!”²

The benefits of living in Britain that the young Erasmus described to Fausto were not just the educational ones he mentioned many years later to the Dutch

¹ ‘Letter to Nicolaas Cannius of Amsterdam, May 29, 1527’, in Erasmus (1960, p. 46).

² This quotation and the following one are from the ‘Letter to Fausto Andrelini, Summer 1499’, in Erasmus (1960, pp. 39-40).

student. First among them was Britain’s “abundance of young girls, attractive and friendly, with the prettiest of faces, who you’d readily value above the Muses, and who bestow kisses upon you at meeting and parting.” To other friends, Erasmus also praised the English food, and the “extraordinarily pleasant and healthy climate”.³ Even in his youth, however, he valued the “delicious” company of English philosophers above all: their learning—“not timeworn and tired, but mature and refined”—their openness and easy conviviality, and, especially, their ability to carry on a debate over dinner or in the pub without falling into a “drunken orgy”.⁴

After five years in London, I have found no reason to disagree with Erasmus about British philosophers. The Philosophy Department at UCL, especially, merits Erasmian terms of praise. The staff’s and graduate students’ love of philosophy, of open discussion, and of a good drink were revealed to me the first day I arrived—the day of the annual graduate conference and dinner—and shaped my time as a graduate student there. The staff/student seminars taught me how to approach philosophical texts: critically, of course, but also much more patiently and carefully than I was inclined to, with respect for the text and a sense that what is good or powerful in it might be subtle or elusive. In spite of their superior learning, the staff made these seminars into a project in which we all felt ourselves equal participants.

Many people made the years at UCL good ones; but four people especially helped make them so. Jo Wolff was an extraordinarily encouraging supervisor: open-minded and interested in my interdisciplinary work and supportive in times when I struggled with my writing. Jo often took the time during supervisory meetings to

³ ‘Letter to Robert Fischer, London, December 5, 1499’, in Erasmus (1960, p. 46).

⁴ Ibid. and ‘Letter to Johannes Sixtinus’, Oxford, November 1499, in Erasmus (1960, p. 40).

discuss a wide range of topics in egalitarian thought, but was also always careful to direct my attention towards the overall structure of my argument.

Véronique Munoz-Dardé impressed upon me, in discussion, and by example, that moral and political philosophy requires much more than cleverness, a delight in solving puzzles or in finding out where important authors went wrong: it requires a realisation of which issues really matter, and, however tentatively held or expressed, a philosophical perspective of one's own.

Mike Otsuka, more than anyone, has taught me how to do analytical philosophy. His careful criticism of my writing, and our long discussions in Central London pubs about a variety of philosophical puzzles and positions showed me the care and precision required for good philosophical argument. The impressive array of counterexamples which he offered to the consequentialist moral thinking to which I was prone when I arrived in London, and the moral and political contractualism to which I am now attracted, have made it all the more difficult to achieve what Véronique has convinced me is so necessary. Should I, one happy day, be able to formulate a moral and political position that both rings true to me and that can withstand his criticism, then I will be pretty sure it will be worth believing.

Ken Binmore has only himself to blame for the failure of his avowed ambition to make me abandon philosophy for something more useful. His supervisions were extraordinarily challenging and helpful, and inspired me to stick to my chosen path. His patient explanation of the central concepts of rational and social choice theory, and his work with me on some philosophical challenges to rational choice theory, greatly contributed to my understanding of these disciplines.⁵

⁵ See Binmore and Voorhoeve (2003) and Voorhoeve and Binmore (forthcoming) for our work on these topics.

My colleagues in the LSE Philosophy Department made LSE a very welcoming place. I owe special thanks Jason Alexander, who took over much of my late Summer grading so I could spend more time on this thesis.

Many others commented on one or several chapters of this thesis. I have acknowledged them by name in the relevant chapter(s).

I also gratefully acknowledge financial support from the AHRB, the Bisschoffsheim Stichting, the Dr. Hendrik Muller's Vaderlandsch Fonds, the Radboudstichting, the UCL Graduate School, the UCL Philosophy Department, the Analysis Trust, and the Mind Association.

Without pressuring me to follow any particular path, my parents offered unshakable support for my various projects over the years. I am deeply grateful to them, and hope in my work as a lecturer to transmit some of their love of learning and discussion, and their awareness of the potential beauty and importance of ideas.

Contents

Introduction	8
1. Welfare	16
2. Equality	36
3. Responsibility	57
4. Equality, Responsibility, and Social Choice	101
5. An Egalitarian Ethos?	140
Bibliography	180
Endnotes to Chapter 4	188

Introduction

In the following chapters, I offer a critical examination and qualified defence of the view that equal opportunity for welfare is the best way of meeting the joint demands of three liberal egalitarian ideals: distributional equality⁶, substantive responsibility⁷, and respect for individuals' differing reasonable judgements of their own good. I also examine which social choice rules best represent the demands of equal opportunity for welfare and the related demand to maximise the opportunities for welfare of the least advantaged. Finally, I examine whether this conception of distributive justice should be pursued by means of the organisation of the major formal institutions of society only, or whether individuals should also adopt it as one of their aims in their everyday life.

⁶ I follow Ronald Dworkin (2000) in using the term 'distributional equality' to denote issues pertaining to the distribution of non-political resources that individuals can use to pursue their particular conceptions of the good life. It contrasts with political equality, since it is neither concerned with the distribution of political power, nor with political rights. It is a disputed question whether distributional equality should be concerned only with resources that individuals can use to pursue their own good, or whether it should also be concerned with people's ability to pursue non-political, non-self-regarding aims. In what follows, I will be concerned only with people's ability to pursue their self-regarding aims. I do not thereby mean to suggest that this is all that distributive justice should be concerned with.

⁷ I follow Thomas Scanlon (1998) in defining substantive responsibility as the way a person's claims on others, others' claims on her, and the quality of her situation should depend on the opportunities she has and the choices she makes. It is distinct from moral responsibility, which is concerned with when it is appropriate to take a person's attitudes or actions as the basis for moral appraisal.

The easiest way to introduce my main arguments is through a short discussion of the views which I take as my point of departure and the debates to which I contribute. In the first three chapters, I have taken Richard Arneson's (1989, 1990a, 1990b, 1990c, 1997) version of equal opportunity for welfare as a starting point for my analysis. Arneson was among the first to argue for an ideal of distributional equality which required that individuals be in a position to achieve equally valuable outcomes through their choices. Distributional equality, in his view, did not require equality of outcome, but instead only the elimination of inequalities for which individuals could not be held responsible. (This has become known as the ideal of 'brute luck equality'.) This understanding of distributional equality opened up room within egalitarian thought for an ideal of responsibility which not only permitted, but, in certain circumstances, required that individuals' achievements depend on their choices. Providing everyone with equal opportunities therefore appeared to offer the possibility of simultaneously meeting the demands of equality and responsibility.

Arneson also argued that respect for individuals' differing conceptions of their own good required that these equal opportunities should be for welfare, conceived of as the degree of satisfaction of the self-regarding preferences a person would after ideally extended and fully informed deliberation. For, he argued, only this conception of welfare would always fully respect each individual's rational judgement of her own good. Moreover, in using this conception of welfare, the state would not be endorsing any particular view of the good life as superior to any other. This view of welfare would therefore allow the state to be appropriately neutral between differing substantive views of the good life.

In the first three chapters, I discuss these conceptions of welfare, equality, and responsibility in turn. Initially, I planned in chapter 1 to offer a qualified defence of a

degree of preference satisfaction conception of welfare. For it seemed to me that Arneson was right to hold that so long as the individual in question was rational and well-informed and had thought the matter over carefully, only certain very significant moral concerns, such as a concern about the quality of the conditions under which her preferences were formed, might permit the state's assessment of her good to depart from her own view of her good. Since the degree of preference satisfaction measure alone among potential conceptions of welfare promised to fully respect each individual's informed, well-considered view of her good, this seemed to me an important reason to try to develop a defensible version of it.

I have come to believe, however, that the degree of preference satisfaction measure cannot fulfil this promise, and chapter 1 is now devoted to arguing for this conclusion. In brief, the argument proceeds as follows. I first point out that people care not just about the satisfaction of whatever preferences they end up having, but also about the values, aims, and attachments, and therefore the preferences that they have. (Someone might care, for example, not just about being a successful artist, but also about having what one might call 'artist's preferences': an appreciation of art, and a desire to be involved in its production.) This means, I argue, that we should take the satisfaction of what I call a person's 'extended preferences'—her preferences over cases in which she is in a particular situation with particular preferences—as our measure of her welfare. I then point out that the possibility of a change in a person's extended preferences makes it impossible for this measure of welfare to always respect a person's view of her good. For, I argue, if we evaluate a potential change in a person's extended preferences from the perspective of the degree to which her future, post-preference change preferences would be satisfied, then we do not respect her current view of which preferences it would be best for her to have. If, by

contrast, we evaluate a potential change in a person's preferences from the perspective of her current extended preferences, then we implausibly disregard the view of her own interests that she would have if the preference change occurred. In sum, whichever method of evaluation we choose, we cannot always respect a person's view of her own good. This conclusion leads me to suggest that we should adopt an alternative conception of welfare, which is based on a list of goods and conditions that are recognised as valuable from the perspective of a variety of different conceptions of the good.

In chapter 2, I offer a qualified defence of the ideal of brute luck equality. I argue that brute luck equality is an important ideal, but that it applies only to situations of 'distributional conflict', in which improving someone's opportunities involves worsening at least one other person's opportunities. In such situations, I argue, each person's equal moral worth gives him a claim to an equally valuable share of resources, where this value is measured in terms of the welfare these resources allow him to achieve. By contrast, in situations where we can improve at least one person's opportunities without worsening anyone else's, we generally do not fail to respect each person's equal moral worth by doing so, even if this leads to inequalities. The one exception to this rule are cases where maintaining brute luck equality is required to symbolically affirm individuals' equal worth in the face of injustices which deny it.

In the same chapter, I also discuss the relationship between brute luck equality and a social and political ideal of equality. Social and political egalitarians stress the need to maintain the social and material conditions which prevent domination, which support individuals' sense of their equal worth as citizens, and which support the attitudes that are required for just social co-operation. Though some have argued that brute luck equality conflicts with the ideal of social and political equality, I argue that

the relationship between the two ideals of equality is better seen as a complementary one. The social and political ideal of equality, by focusing on the relationships between individuals in society and their character and attitudes, identifies important concerns that extend beyond those that are addressed by brute luck egalitarianism. It also has distributive implications, so that only a subset of the distributions that are compatible with brute luck equality may be permissible from the perspective of social and political equality. In turn, the ideal of brute luck equality identifies a way in which inequalities may be objectionable that is missed by the ideal of social and political equality.

In chapter 3, I develop a view of substantive responsibility that is compatible with the ideal of brute luck equality. Following Dworkin (2000, chapter 2), several opportunity-egalitarians have endorsed something like the following idea of responsibility for the economic realm: the choices open to individuals, and their concomitant benefits and burdens, should, as far as possible, be those that would obtain in an Arrow-Debreu type laissez-faire market if individuals entered it with equally valuable bundles (see Eric Rakowski (1991, chapter 3) and Will Kymlicka (2002, pp. 72-73)). This view of responsibility and its relation to brute luck equality have been analysed in detail by Marc Fleurbaey (1994, 1995a, 1995b, 1995c, 1998), amongst others.

Partly because of the difficulties with this view of responsibility that Fleurbaey's analysis has revealed, and partly because of independent doubts about the normative force of the distribution that would obtain in a laissez-faire market in which everyone entered with equally valuable bundles, I choose to approach the issue of responsibility from a different perspective. Through a critical analysis of Scanlon's (1998) work, I develop two views of substantive responsibility, which I call the

Realised Value of Option Set view and the Potential Value of Option Set view. Both views focus on the ways in which giving people certain choices can be valuable for them. The difference between the two views is that, on the Realised Value of Option Set view, the justification of arrangements that give people certain choices is based entirely on how well people end up under these arrangements, whilst under the Potential Value of Option Set view, the justification of such arrangements is based on the value that individuals can achieve (but don't necessarily achieve) through their choices. I argue that the Potential Value of Option Set view is best suited to the ideal of brute luck equality and conclude that we should understand the opportunity-egalitarian ideal as requiring that we provide everyone with the most valuable equal option set, and that this value should be determined by the Potential Value of Option Set view.

In chapter 4, I examine how we can represent this requirement in social choice rules. I also examine the theoretically more difficult question which social choice rule we should use if we replace brute luck equality by absolute priority to those who are most disadvantaged by brute luck.

Two ways of determining who is most disadvantaged by brute luck have been discussed in the responsibility-sensitive egalitarian social choice literature. One way is to take the most disadvantaged to be all individuals who are least well off compared to others who have chosen comparable options (see John Roemer (1993, 1996, 1998, 2002, 2003)). For example, if we are considering returns to education, then we might count the individuals who have the lowest returns to a given educational choice (say, dropping out of high school) as among those who are most disadvantaged by brute luck. For others who chose a similar option (e.g. dropping out of high school) are better off due to no choice of their own, and this is a matter of brute luck. I will call

the approach that aims to improve the situation of the least well off, so defined, the Leximin Value of Options approach.

A second approach starts by assessing the value of individuals' option sets as a whole and then takes those who are most disadvantaged by brute luck to be all individuals with the least valuable option set (see Van de gaer 1993). In the previous example, this would involve first assessing each individual's educational choices and the returns to each of these choices, and then counting those individuals whose overall set of educational options was least valuable as among those who are most disadvantaged by brute luck. I will call the approach that aims to improve the situation of the least well off, so defined, the Leximin Value of Option Set approach.

I argue that under the assumption that individuals can choose among their options under adequate conditions of choice, brute luck inequality in the value of individuals' option sets as a whole is more important than brute luck inequality in the value of individuals' options. I conclude that the Leximin Value of Option Set approach is a better way of integrating priority for those who are disadvantaged by brute luck with the demands of responsibility.

In chapter 5, I ask whether citizens of a just egalitarian society should pursue their egalitarian ideals through the design of equality-promoting formal public institutions alone, or whether they should also sometimes be guided by egalitarian ideals in their choices within the space permitted for free individual choice by these formal public institutions. On first consideration, it may seem natural to assume that egalitarian justice requires both equality-promoting formal public institutions and an egalitarian ethos. For why should the pursuit of justice be limited to formal public institutions and the behaviour required to support these institutions, especially if

individuals' choices and attitudes in their everyday lives can have a profound impact on others' prospects?

However, recent debates on the scope of justice in Rawls's work have thrown up some challenges to this view. I review some of this debate in order to examine what we can learn from it about the roles of formal public institutions and principles for individual conduct in egalitarian justice. I conclude that an egalitarian ethos encompassing both certain principles which Rawls identifies—the natural duties of respect and mutual aid, and the principle of fidelity—and further principles regulating individual conduct and motivation should be seen as part of an egalitarian theory of justice.

1. Welfare⁸

Introduction

Suppose that we agree that for questions of justice in a pluralistic society, we need a public standard of welfare. An appropriate public standard of welfare will have to meet the following two requirements. First, its conception of each person's welfare should, to the greatest reasonable extent, be something that each person can recognise as encompassing the things she wants for herself and as giving these things weights that reflect the relative importance she gives to them. Second, it should be sensitive to the fact that reasonable people hold conflicting conceptions of what constitutes an individual's welfare. It should therefore, to the greatest reasonable extent, respect neutrality of judgement by refraining from endorsing any particular conception of welfare as superior to any other.

In an influential set of essays, Arneson (1990a, 1990b, 1990c) has argued that the following conception of welfare is ideally suited to these requirements: equate each individual's welfare with the degree of satisfaction of her ideally rational, self-

⁸ This chapter will appear as 'Preference Change and Interpersonal Comparisons of Welfare' in *Preference Formation and Well-Being*, edited by Serena Olsaretti, Cambridge: Cambridge University Press. Earlier versions of this chapter were presented at the Royal Institute of Philosophy conference on Preference Formation and Well-Being at St. John's College Cambridge in July 2004, in the LSE Choice Group Seminar in January 2005, and the ECAP 5 Workshop on Philosophy, Economics, and Public Policy in Lisbon in August 2005. I am grateful to those present at these meetings, and especially to Richard Arneson, Luc Bovens, Marc Fleurbaey, Christian List, Andrew Williams, and Jo Wolff for their comments. I also thank Richard Bradley, Michael Otsuka, and Serena Olsaretti for detailed comments on earlier versions of this chapter.

regarding preferences. These are the preferences she would have on behalf of herself if she were to engage in ideally extended deliberation with full pertinent information, in a calm mood, while thinking clearly and making no reasoning errors (see Arneson 1990a, pp. 162-163). (For simplicity, in what follows, I will use the term ‘preferences’, to refer to these ideally rational, self-regarding preferences.)

Arneson argues that this standard of welfare meets the two aforementioned requirements in the best way possible. It meets the first requirement, he argues, because it comes as close as possible to adhering to a person’s own view of her welfare within the constraints set by the need to avoid the intuitively unpalatable move of considering something of value to her that she only considers to be so because of a lack of information or incomplete or erroneous deliberation (1990a, p. 163). It meets the second requirement, he argues, because it does not involve any commitment on the part of the state to a substantive view of what is good for individuals. As Arneson (1990b, p. 450) puts it: “the good in this conception is an empty basket that is filled for each individual according to her considered evaluations.”⁹

In this chapter, I will argue that the fact that people care about which preferences they have, and the fact that people can change their preferences about which preferences it is good for them to have, together undermine this case for accepting a degree of preference satisfaction conception of welfare.

The chapter is organised as follows. In section 1.1, I introduce three concepts of importance to the discussion of a preference satisfaction measure of welfare. First, a person’s *preference type*, which encompasses all the things that determine how she would evaluate, after ideally rational and well-informed deliberation and from the

⁹ See also Otsuka (2003, pp. 110-112).

perspective of her self-interest, her situation and her evaluative dispositions. Second, a person's *limited preferences*, which are her preferences over alternatives in which her situation differs whilst her preference type remains unchanged. Third, a person's *extended preferences*, which encompass her preferences over alternatives in which either her situation, or her preference type, or both, differ.

In section 1.2, I argue that the interest in having the preferences one wants to have is of central importance in human life, and that the degree of preference satisfaction approach should therefore attempt to judge a person's welfare by the degree to which her extended preferences are satisfied.

In section 1.3, I argue that the possibility of a change in a person's extended preferences creates great difficulties for a measure of welfare based on the degree of satisfaction of a person's extended preferences. For if we evaluate a potential change in a person's extended preferences from the perspective of the degree to which her future preferences would be satisfied, then we do not adequately represent each individual's current interest in shaping her future preferences. If, by contrast, we evaluate a potential change in a person's preferences from the perspective of her current extended preferences, then we implausibly disregard the view of her own interests that she would have if the preference change occurred. In sum, it seems that any degree of preference satisfaction measure will have significant drawbacks, since it will involve either neglecting individuals' current interests in shaping their future preferences, or, in their future, judging their welfare by a set of values which might be very alien to them.

In the final section, I argue that this should prompt us to develop alternative measures of welfare. I suggest that one promising candidate is a substantive measure

of welfare based on a list of goods and conditions that are recognised as valuable from the perspective of a variety of different conceptions of welfare.

1.1. Preference-based interpersonal comparisons of welfare

Let us start with a simple description of what is involved in preference-based interpersonal comparisons of welfare. For simplicity, I will limit the discussion to cases where, from the perspective of his self-interest, a person cares only about his own situation, and is indifferent to other people's situations and their attitudes towards him. Let us begin by introducing the notion of a *preference type*. A preference type encompasses all the things that determine how a person would evaluate, after ideally rational and well-informed deliberation and from the perspective of his own self-interest, his situation and his own evaluative dispositions. (From now on, I will drop reference to a person's evaluations being those he would have after ideal and fully informed deliberation and from the perspective of his self-interest, and take them to be so.) A person's preference type therefore tells us how he would rank each combination of his personal situation and evaluative dispositions, and also which evaluative dispositions he has. In a terminology which will shortly be introduced, this is equivalent to saying that two people have the same preference type just in case what I will be calling their 'limited preferences' and their 'extended preferences' over states of the world in which they occupy positions that are the same in all relevant respects are identical.

Let $\{A, B\}$ be the set of preference types consisting of artist's preferences (A) and banker's preferences (B). Let \mathbf{S} be the set of all possible states of the world. A

state of the world is a description of all relevant aspects of each person's situation. Let \mathcal{S} be a state of the world in \mathbf{S} . Let u_i^t be a Von Neumann and Morgenstern utility function defined on the set \mathbf{S} , representing the preferences of a person with preference type t who occupies person i 's position in all possible states of the world. It assigns a real number $u_i^t(\mathcal{S})$ to being in person i 's position in state of the world \mathcal{S} and is bounded both above and below.¹⁰ Because this function only represents a person's preferences over states of the world while keeping his preferences fixed at type t , I will refer to the preferences it represents as a person's 'limited preferences'.

The set $\mathbf{S} \times \{A, B\}$ stands for the set of all pairs (\mathcal{S}, t) with \mathcal{S} in \mathbf{S} and t in $\{A, B\}$. Let v_i^t be a Von Neumann and Morgenstern utility function defined on the set $\mathbf{S} \times \{A, B\}$. The function v_i^t represents the preferences of a person with preference type t over (state of the world, preference type) pairs in which he occupies person i 's position. It assigns a real number $v_i^t(\mathcal{S}, t')$ to each pair (\mathcal{S}, t') in the set $\mathbf{S} \times \{A, B\}$ and is bounded both above and below. Because the function v_i^t represents how a person with preference type t would order a set of options that involve occupying person i 's position in either different states of the world or with

¹⁰ The fact that the utility function is bounded both above and below means that there exists some numbers a and b such that $a \leq u_i^t(\mathcal{S}) \leq b$ for each \mathcal{S} in \mathbf{S} . This means that in no case is being in person i 's position ascribed a utility of negative or positive infinity. This assumption avoids certain decision-theoretic paradoxes that arise when utilities of negative or positive infinity are permitted. See Binmore (1991).

different preference types or both, I will refer to the preferences it represents as this person's 'extended preferences'.

To write that $u_i^t(\mathbf{C}) > u_i^t(\mathbf{D})$ means that, keeping his preference type fixed at type t , a person with preference type t will prefer occupying person i 's position in state of the world \mathbf{C} to occupying person i 's position in state of the world \mathbf{D} . To write that $v_i^t(\mathbf{C}, A) > v_i^t(\mathbf{D}, B)$ means that keeping his preference type fixed at type t , a person with preference type t will prefer occupying person i 's position in state of the world \mathbf{C} with preference type A to occupying person i 's position in state of the world \mathbf{D} with preference type B . By way of illustration, suppose Paul has artist's preferences. Suppose that in \mathbf{C} , Paul is a struggling artist, and in \mathbf{D} , he is a successful banker. Then $u_{Paul}^A(\mathbf{C}) > u_{Paul}^A(\mathbf{D})$ means that, keeping his artist's preferences constant, Paul prefers being a struggling artist to being a successful banker. This preference will be relevant to his choice of career, so long as his choice of career does not change his preference type. By contrast,

$v_{Paul}^A(\mathbf{C}, A) > v_{Paul}^A(\mathbf{D}, B)$ means that Paul prefers being a poor artist with his current artist's preferences to being a rich banker with banker's preferences. This preference would determine, for example, his choice between going to art school (which, let us assume, will maintain his artist's preferences and lead to a career as a struggling artist) and going to business school (which, let us assume, will lead to a preference change to banker's preferences followed by a successful career in banking). We can imagine he holds this preference because he believes that a life devoted to art is superior to one that is not, no matter how successful the latter is. He therefore values having his artist's preferences to such a degree that he would not want to be rid of them and have them substituted by banker's preferences (which, let

us suppose, involve a desire to compete and succeed in the world of high finance and no appreciation of artistic values), even at the cost of being poor and unrecognised in the work he would do as an artist as opposed to wealthy and successful in the career he would choose if he were to acquire banker's preferences.

Both individuals' limited preferences and their extended preferences may differ. Suppose for simplicity that like Paul, Rob is a struggling artist in \mathbf{C} and a successful banker in \mathbf{D} . Suppose, further that Rob has banker's preferences, and that, keeping his banker's preferences constant, this means he would rather be a successful banker than a struggling artist: $u_{Rob}^B(\mathbf{C}) < u_{Rob}^B(\mathbf{D})$. In addition, suppose Rob is committed to the competitive values that underlie his preference for banking, so that he would not accept an opportunity to acquire artist's preferences, especially not at the cost of then having to live as a struggling artist, so that $v_{Rob}^B(\mathbf{C}, A) < v_{Rob}^B(\mathbf{D}, B)$. In sum, in contrast to Paul, Rob believes it is worse to be a struggling artist with artist's preferences than to be a successful banker with banker's preferences.

It is worth noting that it is not necessarily the case that when individuals' preference types differ, both their limited and extended preferences differ. Two individuals with different preference types might have the same limited preferences, but different extended preferences, or the same extended preferences, but different limited preferences.

As an example of the former, consider the case of two gourmands who both enjoy exactly the same dishes to an equal extent: in environments in which they face only choices about what to eat, they will evaluate all options in exactly the same way, so that (at least in these environments) their limited preferences are the same. However, one of the two would prefer, if given the chance, to give up his taste for

fine dining and develop a taste for music instead, so long as he would have an adequate opportunity to enjoy music with his new tastes. The other, by contrast, would not prefer to develop such tastes, so that their extended preferences are different.

As an example of the latter, consider two hedonists, who both rank all (person's position in a state of the world, preference type) pairs in the same way, viz. according to the pleasure they yield, so that their extended preferences are identical. However, one of them likes music, but takes no pleasure in eating, whereas the other takes no pleasure in music, but enjoys a good meal, so that their limited preferences differ.

Let us now turn to preference-based interpersonal comparisons of welfare. A preference-based standard of welfare involves making judgements about whether occupying Paul's position in state of the world \mathbf{C} with preference type \mathcal{A} is better, worse, or just as good as occupying Rob's position in state of the world \mathbf{D} with preference type \mathcal{B} . More precisely, let the function w_i be a Von Neumann and Morgenstern utility function representing this public standard of welfare. The function w_i then assigns a real number $w_i(\mathbf{S}, t)$ to each pair (\mathbf{S}, t) in the set $\mathbf{S} \times \{\mathcal{A}, \mathcal{B}\}$ and is bounded both above and below. This $w_i(\mathbf{S}, t)$ stands for the value of occupying person i 's position in state of the world \mathbf{S} with preference type t . In attempting to determine these values with reference to a person's degree of preference satisfaction, we face two questions. First, whether we should use the degree of satisfaction of a person's limited or extended preferences in determining his welfare. Second, how we should evaluate options that involve preference change. I address the first of these questions in the next section, and the second one in section 1.3.

1.2. Extended preferences and welfare

As mentioned, people typically do not just care about having the world conform to their preferences; they also care about which values, aims, attachments, and therefore preferences they have. This interest in having the preferences one wants to have is, moreover, an important one. Considering people who could be said not to care, or not to care deeply, about their values and aims can illustrate this importance.

Consider first what the life would be like of someone who was completely indifferent about his values and aims. This person's life would be devoid of the kind of commitments and relationships that are a central part of most people's lives. This is evident in cases of commitments to moral ideals: being committed to a cause like eradicating world poverty, for example, involves more than having a particular pattern of desires connected to that cause, such as that it should be realised; it also involves wanting to maintain one's desire for its realisation. But is it also a feature of other commitments that are a central part of people's identity. Being committed to being an artist, for example, does not just involve trying to succeed as an artist, but also to want to maintain and develop one's appreciation of art.

Furthermore, deep friendship involves not just caring about one's friend, enjoying her company, and being ready to help her out when she needs help, but also being prepared to take steps to maintain these attitudes towards her. Similarly, being a loving partner involves not just desiring to share one's life with one's partner, desiring to see him do well, etc. but also to actively maintain and, when necessary,

reinforce these desires (see Frankfurt (2004)).¹¹ In these cases, a person does not just desire to have certain preferences because having them would be instrumental to some other end that she has (eradicating world poverty, being a successful artist, furthering the welfare of one's friend or lover), but also because she regards these as the right preferences for her to have.

More generally, the life of someone who was completely indifferent about his values and aims would be devoid of a particular kind of agency: action directed not merely at shaping his environment to satisfy his desires, but also at shaping himself, in the sense of shaping his values and aims (see Frankfurt (1982, p. 83)). As a consequence, if his life showed any unity of purpose, it would not be the product of any action on his part intended to give his life any particular direction, but rather the product of causes in which he played no active part, or the unintended by-product of his actions.

Similar conclusions apply in the case of a person who, though she has preferences about which preferences she has, ranks her present and potential preferences only on the grounds of the ease with which they can be satisfied (see Dworkin 2000, pp. 292-293). Such a person would also lack the particular attitudes necessary for being substantively committed to any particular cause, relationship, or set of values. As a consequence, she would not purposefully shape her life and herself in accordance with the demands of such commitments.

In sum, the interest in having the preferences one wants to have is essentially the interest in one's ability to shape oneself and one's life in accordance with the demands of the causes, values and relationships to which one is devoted. Given the importance of this interest, we should attempt to base a degree of preference

¹¹ See also Voorhoeve (2003) for a discussion of Harry Frankfurt's views on love.

satisfaction measure of welfare on the satisfaction of a person's extended preferences, since these represent both a person's interests in his situation and his interests in his preferences.

1.3. Preference change and the degree of satisfaction of a person's extended preferences

Let us now turn to the method for determining the degree of satisfaction of a person's extended preferences. Suppose there are four states of the world: one in which Paul is a struggling artist (**C**), one in which he is a successful banker (**D**), one in which he is an unsuccessful banker (**E**), and one in which he is a successful artist (**F**). For any Von Neumann and Morgenstern utility function, we are free to fix the zeros and units.¹² Once we do so, the utilities of all states of the world are fixed. In order to be able to interpret the number $v_i^t(\mathcal{S}, t')$ as the degree to which occupying person i 's position in state of the world \mathcal{S} with preference type t' satisfies the extended preferences of someone with preference type t , we therefore proceed as follows. We set the value of what, from the perspective of type t is the best possible (occupying person i 's position in a state of the world, preference type) pair to one, and the value of the worst pair to zero. For example, suppose that, from the perspective of his current artist's preferences, Paul would consider being a struggling artist with banker's preferences the worst possible pair, and being a

¹² For an introductory discussion of Von Neumann and Morgenstern utility functions and their properties, see Binmore (1991, chapter 3).

successful artist with artist's preferences the best possible pair. We then take

$$v_{Paul}^A(\mathbf{C}, B) = 0, \text{ and } v_{Paul}^A(\mathbf{F}, A) = 1.$$

The value of all other (occupying his position in a state of the world, preference type) pairs will then be determined as follows. We take the number assigned to any other pair to be equal to the probability p that would render Paul indifferent between accepting that pair and a lottery with probability $(1-p)$ of ending up in his position in \mathbf{C} with preference type B and probability p of ending up in his position in \mathbf{F} with preference type A . In this way, each (occupying his position in a state of the world, preference type) pair is assigned a number between zero and one, which we can call the degree to which this pair satisfies Paul's current type A extended preferences over (occupying his position in a state of the world, preference type) pairs. For example, if with artist's preferences Paul would be indifferent between being a struggling artist with artist's preferences and a lottery with a probability of 0.2 of ending up in his position in \mathbf{C} with preference type B and a probability of 0.8 of ending up in his position in \mathbf{F} with preference type A , then $v_{Paul}^A(\mathbf{C}, A) = 0.8$.

Now, we face a difficulty in deciding how to move from the degree to which each pair would satisfy Paul's current extended preferences to an assessment of how well off he would be if each of these pairs were realised. The difficulty is that if we assess each pair by Paul's current, type A extended preferences, this assessment may differ from his own assessment of these pairs once he is in the situation characterised by this pair. For when these pairs involve a preference change to preference type B , then though Paul will now regard this change as making him worse off, once he has preferences of type B , he may regard this change in his preferences as making him better off. For example, from the perspective of his current, artist's preferences, he

might assign the situation in which he is a successful banker with banker's preferences a value of 0.3: $v_{Paul}^A(\mathbf{D}, B) = 0.3$. But if he ended up in this situation through a process of preference change that, from the perspective of his new preferences, he does not regard as in any way a bad one to have undergone, then we may suppose that from the perspective of his new preferences, he would assess this situation as the best possible one: $v_{Paul}^B(\mathbf{D}, B) = 1$. The question is, then, whether we should take Paul's pre-preference change, or post-preference change evaluation as determining his welfare in such cases.

Before we attempt to deal with this question, we should note that in order to assess the impact of a change in a person's extended preferences on his welfare, it is important to assess the conditions under which it takes place. If the preference change was a result of the subversion of Paul's cognitive capacities, or of coercion or oppression, or was a response to an unduly limiting environment, then this might discredit Paul's post-preference change view of his own welfare. I will assume, however, that all preference changes under consideration are not the result of the subversion of a person's cognitive capacities, of coercion, oppression, or unduly limiting circumstances. Preference changes of this kind can occur throughout people's lives; one might, for example, have artist's preferences and through contact with one's friends or one's social environment, or simply through the passage of time, find one's preferences changed to banker's preferences. I will also assume that from the perspective of preference types A and B , having had one's preferences change in this way is not viewed as a bad or a good thing in itself.

Now, the possibility of this change in an individual's evaluation of a particular (occupying his position in a state of the world, preference type) pair means we have

at least two possible ways of judging an individual's welfare by the degree of satisfaction of his extended preferences. The first is to equate the welfare level of each pair with the degree of satisfaction of the extended preferences that he has in that pair. The second is to equate the welfare level of each pair with the degree to which this pair satisfies his current the extended preferences. I will discuss each method in turn.

The first method involves using the extended preferences of type A to evaluate a situation that involves Paul occupying his position in a state of the world with preference type A , and the extended preferences of preference type B to evaluate a situation that involves Paul occupying his position in a state of the world with preference type B . This would mean taking $w_{Paul}(\mathbf{C}, A)$ to be equivalent to $v_{Paul}^A(\mathbf{C}, A)$ and $w_{Paul}(\mathbf{C}, B)$ to be equivalent to $v_{Paul}^B(\mathbf{C}, B)$, and so on.

Doing so means that at every point in time, our standard of welfare will agree with each individual's own assessment of his welfare at that time. Moreover, this standard of welfare will always respect each individual's preferences over options that do not involve changes in his extended preferences. However, this standard will not always agree with an individual's pre-preference change assessment of the value of options that involve a change in his extended preferences. For this measure will count a change from a situation in which Paul is a struggling artist with artist's preferences to a situation in which he is a successful banker with banker's preferences as an improvement in Paul's welfare, since the degree of satisfaction of the extended preferences he has after the change is larger than the degree of satisfaction of his extended preferences before the change:

$w_{Paul}(\mathbf{C}, A) = v_{Paul}^A(\mathbf{C}, A) = 0.8 < w_{Paul}(\mathbf{D}, B) = v_{Paul}^B(\mathbf{D}, B) = 1$. But from the perspective of his current, artist's preferences, Paul will disagree with this judgement.

It follows that this measure does not adequately represent Paul's interest in having the preferences he wants to have: it will not consider the goods and conditions that enable him to sustain his preferences, or develop them in the direction he wants, as contributing to his welfare, unless his sustaining or developing these preferences will contribute to a higher degree of satisfaction of whatever preferences he ends up having. For example, so long as Paul has artist's preferences, this measure will regard the resources and conditions that help him sustain his artist's preferences as of less value to him than the resources and conditions that would lead him to develop banker's preferences, when the latter could be more easily satisfied. Given the importance of the interest in shaping one's tastes, values and aims in the direction one wants, this represents a severe drawback of this version of the degree of satisfaction measure of welfare.

The second method assesses each (occupying his position in a state of the world, preference type) pair from the perspective of his current preferences. To illustrate this method, suppose again that Paul's current extended preferences are those that go with type A . We regard these extended preferences as determining the welfare of all future (occupying Paul's position in a state of the world, preference type) pairs. We would then take $w_{Paul}(\mathbf{C}, A)$ to be equivalent to $v_{Paul}^A(\mathbf{C}, A)$, $w_{Paul}(\mathbf{C}, B)$ to be equivalent to $v_{Paul}^A(\mathbf{C}, B)$, and so on.

This method obviously represents Paul's current interests in his future preferences. However, it does so at the cost of not always respecting Paul's post-preference change extended preferences. Suppose Paul's preferences at time 0 are

artist's preferences. Suppose further that we take the degree of satisfaction of his extended preferences at time 0 as the measure of his current and future welfare. Then we will evaluate a change from a situation at time 0 in which he is a struggling artist to a situation in which at time 1 he is a successful banker as making him worse off. As noted, after this change, Paul will disagree: at time 1 he will regard his new situation as the best possible one. Now, suppose this change does take place, and at time 1 we can present Paul with an opportunity to change his preferences back to artist's preferences at time 2. From the perspective of Paul's extended preferences at time 0, this would be an opportunity to increase his welfare. But from the perspective of his extended preferences at time 1, taking this opportunity would make him worse off. By making his preferences as time 0 normative throughout these periods, we would not be respecting his judgements at time 1. We thus can represent Paul's interests in his future preferences at time 0 only at the cost of disregarding his extended preferences at time 1.

Now, in some cases of preference change, we might have reason to regard a person's initial preferences in sequences of this kind as normative; an example might be a case in which the preference change between time 0 to time 1 was a result of the subversion of Paul's cognitive capacities, or of coercion or oppression, or was a response to an unduly limiting environment. But we have assumed that the process of preference change was not of this sort. In such cases, it does not appear appropriate to judge Paul's welfare at time 1 from the perspective of his very different extended preferences at time 0: this would be judging his welfare by a set of values that he no longer holds.

Though I cannot discuss all possible methods of dealing with the case of preference change that fall within the family of possible degree of preference

satisfaction measures, it seems that all possibilities that involve a compromise between these two approaches will suffer from some combination of the drawbacks of these two methods. For example, consider determining a person's welfare in a given period by the degree of satisfaction of a weighted average of the extended preferences he has over that period, with the weights determined by the relative amount of time he holds certain preferences.¹³ This would involve both limiting the degree to which a standard of welfare represents a person's interest in his future preferences, and assessing his welfare at some points in time by a set of values which he no longer holds. It would, for example, imply that if Paul had artist's preferences for 30 years, and then developed banker's preferences later in life, then (re)developing his taste for art by enrolling in evening classes of art appreciation would improve his welfare even if he had his banker's preferences for 10 years, and saw no value at all in taking these classes.

In sum, it seems that any degree of preference satisfaction measure will have significant drawbacks, since it will involve either neglecting individuals' current interests in shaping their future preferences, or, in their future, judging their welfare by a set of values which might be very alien to them.¹⁴

¹³ Something akin to this possibility, though without the use of the distinction between limited and extended preferences, is discussed by Richard Brandt (1979, pp. 247-253) and Arneson (1990a, pp. 162-167). See also the following footnote.

¹⁴ It may be of interest to note how my discussion of the difficulties which preference change creates for a degree of preference satisfaction measure differs from Brandt's. Brandt (1979) also argues that the possibility of preference change undermines the case for a degree of preference satisfaction measure of welfare. His discussion differs from mine, however, in not making use of the distinction between a person's limited and extended preferences. This distinction is, I believe, crucial to the problem. For not just any change in a person's preferences is generates a problem for a degree of

1.4. A substantive conception of welfare?

This conclusion should, I believe, prompt us to consider alternative measures of welfare. In closing, I would like to outline one approach that strikes me as worth pursuing. This is to construct a public conception of welfare from a list of goods and conditions that can be recognised by people with divergent values as generally important constituents of a good life (see also Scanlon 1991). Some of its constituent elements will be broad categories that can be realised in different ways by people with different values, such as developing one's capacities, leading the life one wants with family and friends, job satisfaction, and achieving success in one's main endeavours. It will also include more specific goods like health, leisure, and wealth, which are generally judged to be important elements of a good life. Our discussion

preference satisfaction metric. A change in a person's limited preferences is not sufficient to undermine the degree of preference satisfaction measure; what is required is a change in a person's extended preferences. To see this, consider the case of a hedonist, who ranks all (occupying his position in a state of the world, preference type) pairs on the basis of the pleasure they yield for her. Suppose that, at time 0, she wants to celebrate her birthday at time 2 with a dinner in her favourite fish restaurant rather than in a steak house, since her current limited preferences are for fish over meat, and she does not expect to undergo a preference change between now and time 2. However, suppose she does undergo a change in her limited preferences, so that at time 2 she prefers to dine in a steak house. So long as this is a change in her limited preferences only, we have no difficulty assessing the welfare of these two options. For we can assess the welfare associated with the four options (eating fish at time 2 with a taste for fish), (eating meat at time 2 with a taste for fish), (eating fish at time 2 with a taste for meat), and (eating meat at time 2 with a taste for meat) from the perspective of her unchanged extended preferences, that is to say, by equating the welfare of each option with the pleasure it yields.

also teaches us that this list should include goods and conditions that generally enable people to maintain or develop the tastes, values, and aims they want to have.

By including categories, goods and conditions that are valuable from the perspective of different views of the good life, this list accommodates both the diverse views of the good life that are held within a population, and the different views of the good life that a person may have during his lifetime. The standard objection to a conception of welfare of this kind is that by using the same list of goods and conditions to assess everyone's welfare, and by using the same weights for these goods and conditions in each person's case, it fails to fully respect each person's view of their own welfare. The force of this objection depends on the assumption that there is an alternative measure of welfare—the degree of preference satisfaction measure—which *does* fully respect each person's view of their own good. But our discussion makes clear that given the possibility of change in a person's extended preferences, no form of the degree of preference satisfaction measure can fully respect each individual's judgements of her own welfare. It may be, therefore, that a substantive conception of welfare of this kind meets our first requirement (to respect, to the greatest reasonable extent, each person's view of their own welfare) because this just *is* the furthest we can go towards respecting each person's view of their own good.

A standard of welfare of this kind also respects our second requirement of neutrality of judgement (see Scanlon 1991, pp. 39-40 and Otsuka 2003, pp. 110-112). For, in attempting to accommodate to the greatest extent possible different views of the good life, it is constructed without the assumption that any particular conception of the good or set of conceptions of the good is the right one to the exclusion of others. In sum, the difficulties that the possibility of change in a person's extended

preferences causes for a degree of preference satisfaction conception of welfare
render a substantive conception of welfare more attractive.

2. Equality¹⁵

Introduction

My purpose in this chapter is to discuss the following three aspects of the ideal of brute luck equality. First, its distributional implications; second, which view of the nature of individuals' claims support it; and third, how its demands relate to a distinct social and political ideal of equality.

The chapter is organised as follows. In section 2.1, I build on Arneson (1989, 1997) to argue that a strict interpretation of brute luck equality requires providing everyone with option sets which meet a demanding set of conditions. I will call option sets that meet these conditions 'equal option sets'. I also suggest that we may adopt a less demanding form of brute luck equality, which requires giving everyone equally valuable option sets.

In section 2.2, I argue that the central brute luck egalitarian idea is a claim about the fair distribution of resources among adults with certain rational capacities in situations in which (a) no one could reasonably be held responsible for the value of the resources available; (b) no one is more deserving of benefits from these resources than anyone else, and (c) there is distributive conflict, in the sense that giving one person more means that someone else ends up with less. The claim is that in such cases, everyone has a *prima facie* claim to an equally valuable share, where this value is measured in terms of what these resources enable people to achieve. This claim is based on individuals' equal moral worth, which in turn is based on their possession

¹⁵ I thank Michael Otsuka and Jonathan Wolff for comments on this chapter.

of a moral sense, an ability to rationally form and pursue a plan of life, and an ability, given adequate circumstances of choice, to make choices for which they can be held responsible. I contrast this claim with Larry Temkin's well-known characterisation of the central brute luck egalitarian idea as the claim that "it is bad, or objectionable, to *some* extent—because unfair—for some to be worse off than others through no fault or choice of their own" (2002, pp. 129-130; emphasis in original). I also discuss its relation to the levelling-down objection.

In section 2.3, I examine the relationship between brute luck equality and a social and political ideal of equality. Though some have argued that brute luck equality conflicts with the ideal of social and political equality, I argue that the relationship between the two ideals of equality is better seen as a complementary one. The social and political ideal of equality, by focusing on the relationships between individuals in society, and their character and attitudes, identifies important concerns that extend beyond those that are addressed by brute luck egalitarianism. It also has distributive implications, so that only a subset of the distributions that are compatible with brute luck equality may be permissible from the perspective of social and political equality. The ideal of brute luck equality, in turn, identifies a way in which distributive inequalities may be objectionable that is missed by the ideal of social and political equality.

2.1 Equal option sets and equally valuable option sets

Brute luck equality requires that any inequality of outcome that obtains is wholly the result of choices for which individuals can be held responsible. The distributive

principle that this ideal entails is best grasped by starting with the following simple case. Suppose each member of a group of individuals has to make only one choice out of a set of options that is available to him. All outcomes of these options are certain. Furthermore, all members of this group have equal and adequate opportunity to inform themselves about the consequences of their choice and to deliberate about it. Finally, their reasoning abilities, will power, and character are such that they have an equal ability to 'negotiate' their options. In this case, brute luck equality obtains among this group of people when (i) each of their option sets contain the same number of options n ; and (ii) their 1st best, 2nd best, ..., and n th best options yield equal welfare (see also Arneson 1989, pp. 85-86). For it will then be the case that any inequality in welfare that results from individuals' choices from their option sets is entirely due to choices for which individuals are responsible.

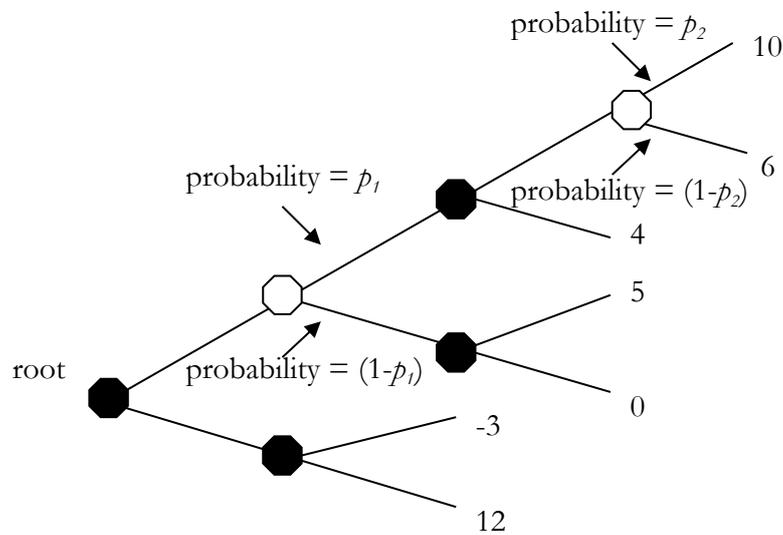
Let us now abandon this simple case and introduce consecutive choices and the role of chance. To discuss such cases, Arneson (1989) uses the helpful device of describing individuals' option sets as decision trees (see also Roemer (1996, pp. 263-272)). Each decision tree consists of a set of paths, each beginning at the root (the onset of the period from which the individual can be held responsible for his actions) and ending at the terminus, when the individual's life is 'played out'. The forking paths which branch out from the root represent different possibilities for that individual. Some forks represent different choices the individual might make. Others represent different possibilities for an individual that are the result of things other than his choices such as the vicissitudes of nature or the choices of others, both of

which I shall refer to as ‘chance’.¹⁶ Paths are of different kinds. A ‘pure choice path’ is a sequence of choices only. A ‘mixed path’ is a sequence of choices and chance occurrences. Figure 2.1 represents a simple decision tree. At white nodes, the path one follows along the subsequent fork is determined by chance. At black nodes, the path one travels along the subsequent fork is determined by choice. The final value of a path is determined at the terminus.¹⁷ The number at the end of each path represents the overall welfare associated with that path as given by the public conception of welfare. The five topmost paths are mixed paths, and the lowest two paths are pure choice paths.

¹⁶ Arneson makes only one reference, in a footnote, to others’ choices as impacting on one’s welfare. In this footnote, he argues that we compute the value of a branch with reference to “how other people are actually expected to behave” (1997, p. 242n.1).

¹⁷ This does not mean that all preference satisfaction comes at the end of one’s life. It merely means that a person’s lifetime welfare if he were to follow a particular path can only be assessed at the terminus of that path. Supposing for simplicity that the welfare generated by two subsequent choices is additive, we might imagine the welfare associated with the bottom path in figure 2.1 to be 5 after his first choice and 7 after the second. This is important for the exercise of imputing the value of a node from the value of the choice set that follows it. For the value of the second node on this bottom path would then have to be computed as 5 + the value of a choice set where one choice yields –8 and another yields 7, and not as the value of a choice set where one choice yields –3 and another yields 12, as figure 2.1 might be taken to suggest.

Figure 2.1. A decision tree



Now, brute luck equality requires that all risk is voluntarily incurred, so that the results of such voluntary gambles can be classified as “option luck”, that is, luck that results from a deliberate and calculated gamble that the person in question should have anticipated and might reasonably have declined.¹⁸ Individuals’ decision trees must therefore include a pure choice path that is a reasonable alternative to all mixed paths.¹⁹ Once this condition is met, brute luck equality will obtain among a group of individuals when all members of this group have equal and adequate opportunity to inform themselves about the consequences of and to deliberate about their choice,

¹⁸ This definition of option luck follows Dworkin (2000, p. 73), with the addition of the word “reasonably”. The reason for this insertion is explained in the next footnote.

¹⁹ This formulation, which I owe to Otsuka (2004), solves the problem of non-voluntarily incurred risk which plagued Arneson’s (1989) first formulation of equal opportunity for welfare. For criticism of this initial formulation, see Lippert-Rasmussen (1999). In response to this criticism, Arneson (1999, p. 491) introduced the requirement that the secure option should be “satisfactory.” This requirement is criticised in Lippert-Rasmussen (2001, pp. 572-573). Otsuka’s formulation is meant to address this criticism as well.

have an equal and adequate ability to negotiate their options, and face identical decision trees. But, at least so long as the value of risky prospects is given entirely by their expected value, it will also obtain among such individuals if the following, slightly weaker conditions are satisfied: (a) all individuals face the same pure choice paths; (b) if we ‘cut off’ individuals’ mixed paths at the last choice node, and assign any remaining risky prospects their expected value, then all individuals face the same ‘truncated’ mixed paths; and (c) all individuals’ option sets include a pure choice path that is a reasonable alternative to the mixed paths. I will call circumstances that meet these conditions circumstances in which individuals have equal option sets.

Might brute luck equality also obtain under less demanding conditions, when individuals’ decision trees differ in the number of branches, or the welfare that these branches yield, or individuals differ in their negotiating abilities? Peter Vallentyne (2002) believes that it might. He argues that one should regard the value of an individual’s option set as determined by the different levels of welfare she can achieve through her choices from that set, as well as by her ability to negotiate her options, and that it is sufficient for brute luck equality to obtain that (a) all individuals’ option sets are equally valuable, and (b) all individuals’ option sets include a secure option that is a reasonable alternative to the risky options. (Given his assumptions, equal option sets will be of equal value; providing everyone with equal option sets will therefore be sufficient, but not necessary to achieve brute luck equality.)

Strictly speaking, providing people with equally valuable option sets is not sufficient for brute luck equality. For providing everyone with equally valuable option sets may be compatible with inequalities in welfare between individuals that are not entirely due to choices for which they can be held responsible. Suppose, for example,

that two individuals' option sets contain two options, with the following payoffs. *A*'s first-best option yields an excellent level of welfare, while *B*'s first-best option yields a very good level of welfare, and *A*'s second-best option yields a low level of welfare, while *B*'s second-best option yields a moderate level of welfare. Suppose that these differences, coupled with *A*'s and *B*'s ability to negotiate their options, are such that by an appropriate measure of the value of their option sets, their option sets are equally valuable. Now, if they both make in some sense the 'same', equally prudent choice, then *B* will be worse off than *A* due to no choice or fault of his own, so that brute luck equality is violated.

Nonetheless, it is also true that in this case, we can say of *B* that he was advantaged by brute luck in a way that *A* was not. For example, if both person's second-best options involved the possibility of making a tempting, self-regarding, but imprudent choice (like indulging in an unhealthy life-style), then *B*, but not *A*, would have faced less bad consequences than *A* if he had made this choice. This counts as a way in which *B* was advantaged over *A*, because, let us suppose, in this case, as so often in human life, it was good to enjoy some protection against the consequences of making an imprudent choice. Thus, though it seems to me that the brute luck egalitarian ideal is only completely realised when individuals have equal option sets, providing individuals with equally valuable option sets may be an acceptable weaker version of this ideal.²⁰

²⁰ See also chapter 4, where I argue that when we are concerned with maximally improving the situation of those who are most disadvantaged by brute luck, we should maximise the value of the least valuable option set.

2.2 The reasons supporting brute luck egalitarianism

The best-known characterisation of the central idea underlying brute luck egalitarianism is Temkin's claim that "it is bad, or objectionable, to *some* extent—because unfair—for some to be worse off than others through no fault or choice of their own" (2002, pp. 129-130; emphasis in original; see also Temkin 1993). This characterisation suggests that brute luck egalitarians see brute luck equality as an intrinsically valuable aspect of a state of affairs, as one among the possibly many considerations which determine the goodness or badness of that state of affairs. A common response to this idea is to argue that this form of value is "mysterious", in contrast with readily intelligible values like the welfare that individuals' enjoy in a particular state of affairs (see, for example, Munoz-Dardé 2005, pp. 272-275). This supposed strangeness of attaching intrinsic value to brute luck equality is often pressed by considering a situation in which the only way to realise equality is through making some individuals worse off, and nobody better off, and asking a (potential) brute luck egalitarian whether in this situation he can really find a respect in which achieving equality through levelling down is good. If, on reflection, he believes that levelling down is not good in any respect, then, it is claimed, he must abandon brute luck egalitarianism.

The contemplation of cases of levelling down has indeed led some onetime brute luck egalitarians to abandon their commitment to this ideal. Arneson (1999, pp. 232-233), for example, writes:

"I find plausible the bare claim that equality, and more generally, how one person's condition compares to that of others, do not matter either morally or intrinsically. When I contemplate cases of levelling down, in which

equality can be achieved by destroying the advantages now possessed by better-off persons without in any way improving the condition of anyone else, I do not judge that there is one respect in which the outcome of levelling down is an improvement—it creates equality—even though perhaps, all things considered, the change is not morally desirable. The levelling down seems a waste, pure and simple, and everyone's having the same does not seem in any way intrinsically worthwhile.”

But of course, a brute luck egalitarian can also respond that achieving brute luck equality *is* one respect in which levelling down is good, because it makes things fairer, but that this respect may be outweighed by other considerations, so that the best thing overall would be not to level down. This is indeed how many brute luck egalitarians have responded to cases of levelling down (see Temkin 2002, p. 155; 1993, p. 282; and Otsuka 2004).

In sum, a common exchange of arguments leads to one of two positions: either one does not see anything good in achieving brute luck equality through levelling down, and so abandons one's belief in brute luck equality, or one affirms that there is something good in achieving brute luck equality by levelling down (while judging that this good may be outweighed by the bad occasioned by some individuals' welfare loss), thereby affirming one's attachment to the brute luck egalitarian ideal.

In what follows, I want to defend a view that is distinct from both of these positions: I want to maintain that brute luck equality is an important ideal, whilst simultaneously maintaining that, in certain cases, there is no respect in which levelling down is good.²¹

²¹ Tungodden (2003, p. 9) notes the possibility of this position.

Let me first turn to the second part of this claim. To begin with, we should note that it is not true that achieving equality through levelling down is never intrinsically valuable in any way. In certain cases, for example, one may be moved to level down by a sense of solidarity with those who are worse off. Feelings of this kind are easily understandable in cases when the person who is better placed shares an important end or aim with the person who is less well placed, or where he identifies strongly with the less well placed person, and in which as a consequence of levelling down, the content of his experience will be similar in relevant respects to the originally less well placed person. In Roman Polanski's film *The Pianist*, for example, the main character, Wladyslaw Szpilman, escapes from the Warsaw ghetto with the help of members of the Jewish resistance, who themselves stay behind to prepare their uprising against the Germans. He watches from an apartment in which he is hiding as the ghetto uprising begins and is subsequently crushed by Nazi troops. At this moment, Szpilman is pained by the thought that he should be with the inhabitants of the ghetto, to fight (and probably to die) with them. He knows that staying would have done no one any good: he was weak, and ill-suited to killing. Moreover, his escape was not a betrayal of those that stayed behind: they helped him escape, and, like many others who risked their lives for him during the War, wanted a man of his talents to survive. We may suppose that later on, he comes to believe that it was not wrong, all things considered, for him to escape. Nonetheless, forgoing the opportunity to escape and therefore remaining worse off than he could be, without improving anyone else's situation, would not have been senseless; and his escape, even if morally justified, comes at a moral cost.²²

²² One should note that Szpilman's case is not, strictly speaking, a case of levelling down, since staying behind would mean one person forsaking an advantage that he does not yet enjoy with no benefit to

The type of cases in which solidarity gives one reason to level down appears limited by the following considerations. First, solidarity only gives people a reason to level down in order to share each other's fate when there is unity in the aims and commitments of the people involved, and a shared sense of membership of a group, a membership which is, moreover, an important part of one's identity. Second, it is not enough that the levels of welfare of the individuals in question will be the same after levelling down (and it may not even always be necessary); what is important is that the content of the experience of all individuals involved will be relevantly similar.

The reasons of solidarity, however, do not exhaust the reasons to level down. Jonathan Wolff (2001), for example, argues that levelling down may be intrinsically valuable in cases where equality has the expressive function of affirming citizens' equal worth in the face of injustices that deny this equal worth. As a mayor of a town in the segregationist South, for example, one might have good reason to close the local swimming pool if the only alternative was opening it to whites only, even if opening or closing the pool did not affect the welfare of blacks in any way.

Levelling down might also be valuable because it maintains various valuable social relationships. Richard Norman (1997, p. 252), for example, has suggested that members of a community might have a reason to refrain from a social change that will make all of them better off but which will introduce significant inequalities, because they fear that this might undermine the attitudes that are constitutive of

others for the sake of equality rather than taking away from someone an advantage that he already enjoys with no benefit to anyone else for the sake of equality. But the moral sentiment of solidarity that it illustrates can equally motivate one to see at least something good in levelling down. For we can easily imagine a Jew who had, through good fortune, remained undiscovered in hiding outside the ghetto, feeling that his place was with those inside.

egalitarian social relations. They might fear that with greater inequality, they would “no longer be united by a shared experience and a shared condition” (Norman 1997, p. 252). They might also fear that the more prosperous would become “disdainful and supercilious, and the less prosperous (...) servile [and] resentful” (Norman 1997, p. 252).²³

There are cases, however, in which Arneson’s claim that levelling down is a “waste, pure and simple” strikes me as correct. Imagine, for example, that after a shipwreck, two men, who are strangers to each other, are washed up on two adjacent deserted tropical islands. The men are of equal health and ability, and, if neither of them possessed any tools, their islands would afford each an equal and adequate ability to live decently. The sea between these islands is swept by strong currents, which make it impossible for the two men to travel between them. They are, moreover, so far apart that one cannot discern anything but the other island’s outline with the naked eye. One of the men, though, has had the good fortune of being washed up beside a small box, containing a knife, sunglasses, and a pair of binoculars. With these binoculars, he can observe the man on the other island, though he remains unseen. He notices that his fellow survivor was less fortunate than he, in not having found any such tools or conveniences. Does he have any reason to destroy the possessions he has stumbled upon, in order to achieve equality with the man on the other island? I do not think so. In this case, none of the aforementioned reasons for levelling down apply, and it would seem to me entirely senseless for the more

²³ As Wolff (2001) points out, however, one might argue that this is not really a case of levelling down, since one could include these social relations, and the attitudes constitutive of them, in one’s understanding of individuals’ welfare.

fortunate man to destroy these possessions in order to achieve equality between them.

This does not mean, however, that I think brute luck equality is not an ideal with moral force. Should it become possible, for example, by a change of currents between the islands, for the more fortunate man to send over some of his undeserved benefits at moderate cost—say, by placing the sunglasses on a small raft which he knows will wash up on the other island—then he should do so to the point of achieving equality between them.

It therefore appears to me that brute luck equality is a principle that has moral force in some contexts, but not in others. More specifically, I believe it has moral force in situations of ‘distributional conflict’—situations in which making one person better off involves making at least one other person worse off—but no moral force in situations in which one person or several people can be made better off at no cost to anyone else. (By using the term distributional conflict I do not mean to suggest that there is any actual conflict between individuals involved, just that their interests in furthering their own welfare conflict.) Assuming I am right that brute luck equality does not always have moral force, this would not be something particular to equality. As Kamm (1996) has argued, it is true of many moral principles that they have moral force in some settings, but lack moral significance in others. Furthermore, regarding brute luck inequality as bad only in cases of distributional conflict does not rob it of its teeth. As Tungodden (2003) demonstrates, it retains potentially tremendous force in ranking distributions in such cases.

Let me try to say a bit more in defence of the claim that brute luck equality has moral force only in cases of distributional conflict. I will first argue for the significance of brute luck equality in cases of distributional conflict. Consider a group

of individuals who all have a certain minimum level of the rational capacities required to have a moral sense, the ability to form and pursue a view of the good life, and the ability, given adequate circumstances of choice (including adequate time to deliberate, and adequate information), to freely choose between various courses of action.

Imagine that these individuals face a problem of division of resources in a situation of distributional conflict, in circumstances in which none of them could reasonably have been expected to influence the value of the resources available, and in which no one is more deserving than anyone else of the benefits these resources bring. Then it seems that everyone has a strong *prima facie* claim to an equally valuable share of resources, where the value of these resources is given by the value of the option sets they enable people to choose from. The reason everyone has a claim of this sort is that this is the appropriate way to recognise the fact that each person is of equal worth, because her rational capacities are above a certain minimum level. (This idea of equal worth is therefore based on what Rawls (in TJ, section 77) calls a “range property”. Just as all point within a circle equally have the property of being in the circle, all individuals who have rational capacities within a certain range—above the minimum required for them to have a moral sense and an ability to form and pursue a plan of the good life—are of equal worth.) Moreover, the fact that she is capable of choosing between various views of the good life and courses of action makes it the case that the value of her option set, rather than the value she realises through her choices, is the appropriate way of valuing her share of resources. (The views of individuals’ capacities to choose freely and of their responsibility for their choices that best fit these ideas are discussed in more detail in chapter 3.)

This judgement is, I believe, familiar and widely shared. Once we describe such cases—a band of travellers accidentally shipwrecked on a desert island to which no

one has a prior claim and none of whom is more deserving than another, a group of hikers chancing on a bush of berries in the wild and having to decide how the harvest will be apportioned, or, more abstractly, the claims on worldly resources of people who come into the world equally undeserving—equal division immediately suggests itself as the *prima facie* correct answer so long as we assume that these are cases of distributional conflict (see also Rakowski 1991, pp. 65-74).

What about cases in which there is no distributive conflict, so that we can improve the welfare of at least one person without decreasing the welfare of anyone else? In general, the positive worth of human beings requires us to further their welfare, so that we should improve the welfare of the individual(s) in question. As our previous discussion illustrates, this general reason may in specific cases conflict with the demands of solidarity, the symbolic affirmation of equal worth in the face of injustice, or the desire to maintain certain social relationships, all of which may make levelling down good, at least in one respect. But is there also a moral loss involved in departing from equality to improve some individuals' welfare when these reasons for levelling down do not apply, because individuals' equal worth gives them a *prima facie* claim to an equal share, a claim that will then not be met? I do not think so. It seems to me that in such cases, individuals' equal moral worth is fully respected by the fact that we accept that if the benefits in question could be redistributed amongst everyone, then everyone would have a claim to an equally valuable share. Thus, in our example of the two men shipwrecked on adjacent islands, it seems to me that all that is required for the first individual to fully respect the other survivor's equal moral worth is that he accepts the duty that should it be possible, at moderate cost, to redistribute his good fortune to the point of equality, then he should do so. There is, in my view, no sense in which his recognition of the other's equal moral worth

requires him to destroy these advantages, should this be the only way of achieving brute luck equality.

This means that Temkin's characterisation of the central brute luck egalitarian claim is correct for cases of distributional conflict. In such cases, equal distributions are good because they are fair, and they are fair because they satisfy each individual's claim to an equally valuable share, a claim which is based in individuals' equal moral worth. As mentioned, the idea that individuals have a claim to an equally valuable share in cases of distributional conflict seems to be a familiar and widely held judgement, so that there is no need to posit some mysterious value to explain why we should favour equality.

However, Temkin's claim is incorrect for cases in which we can improve the situation of at least one person without making anybody worse off. Inequalities generated by improving people's situation in such cases are not bad, because they are not unfair. They are not unfair, because individuals' equal moral worth does not require us to maintain equality in such cases; it requires only that we recognise that should it become possible to redistribute these benefits in such a way that we can achieve equality, then we have a significant moral reason to do so.

2.3 Brute luck equality and social and political equality

We cannot explain all the ways in which inequality is bad with reference to the ideal of brute luck equality. There is a long egalitarian tradition that is concerned with the attitudes and material conditions that are required for members of a certain society to

live as equal citizens and to maintain just social co-operation over time.²⁴ From the perspective of this tradition, material and social inequalities are bad when they undermine individuals' ability to live as equal citizens and lead to attitudes which are inconsistent with just social co-operation.

More specifically, this tradition identifies three principal ways in which inequality can be bad. The first is that inequality may lead to domination. Even in the face of equal political liberties, for example, a concentration of wealth and economic power in the hands of a relatively small section of society may enable this section of society to exercise undue political influence. But the domination in question can extend beyond the political sphere. Inequalities between the rich and the very needy may equally give the former great control over the lives of the latter in various ways, especially in the absence of effective legal constraints regulating their interactions.²⁵

Second, inequalities may lead to unwarranted feelings of inferiority on the part of the least well off. For example, a poor person might feel ashamed and inferior because he lacks the resources to participate fully and with dignity in the social life of his society. Furthermore, he may find it difficult not to let his judgement and the judgement of his peers that the majority is better off than he is in many important respects, such as social status, recognised achievement, income, health, leisure, friendships, etc., affect his sense of self-worth.

Third, significant social and economic inequalities may encourage great vices and collectively harmful attitudes in both the less well off and the better off. Large

²⁴ Classic examples are Rousseau (1988), Tawney (1964) and Rawls's work. More recent writers in this tradition include Phillips (1999), Anderson (1999), and Scheffler (2004).

²⁵ The relationship between the construction firm owner and his employees recounted in Robert Tresselt's *The Ragged-Trousered Philanthropists*, for example, vividly illustrates this point.

inequalities may lead the less fortunate to be viewed by others as inferior, which together with their own judgement of their inferiority may lead to widespread attitudes of deference and servility on the side of the less well off, and arrogance and the will to dominate on the side of the better off. Once such ideas of the moral inferiority of the less well off are established, the better off may cease to view others as equal citizens to whom justification of a society's institutions and practices is owed. Richard Tawney made this point eloquently:

“One of the (...) effects of extreme inequality is its tendency to weaken the capacity for impartial judgement. It pads the lives of its beneficiaries with a soft down of consideration, while relieving them of the vulgar necessity of justifying their pretensions. (...) It disposes them, on the one hand, to take for granted themselves and their own advantages, as though there were nothing in the latter which could possibly need explanation, and, on the other hand, to be critical to claims to similar advantages advanced by their neighbours who do not yet possess them. It causes them, in short, to apply different standards to different sections of the community, as if it were uncertain whether all of them are human in the same sense as themselves” (1964, pp. 37-38).

Furthermore, the less well off may become envious and resentful of the advantages enjoyed by the better off, leading them to prefer a social world in which the advantages of the better off are diminished, even if this is not beneficial to them. Awareness of these attitudes may lead to a jealous guarding of their advantages by the better off, and a desire to see their own relative position maintained, even if this comes at a cost to themselves or to the worst off (see TJ, pp. 467-468). Now, social co-operation requires a willingness on the part of all those involved to propose and

abide by fair and mutually beneficial terms of co-operation. The sense on the part of the well off that one need not justify one's actions or social institutions to the less well off that Tawney identifies, and the envy and jealous guarding of relative advantage that Rawls describes, therefore make social co-operation impossible.

Since domination, an unwarranted sense of inferiority, and the vices and harmful attitudes that may accompany large inequalities are objectionable whether or not these inequalities are the result of people's voluntary choices, the ideal of social and political equality will rule out some inequalities that are consistent with brute luck egalitarianism. For its part, brute luck egalitarianism rules out certain inequalities that are consistent with social and political equality. For the latter ideal does not regard inequality as bad so long as it does not undermine equal citizenship among those who abide by the terms of social co-operation. It does not pronounce, for example, on the justice or injustice of inequalities between people not involved in social co-operation within the same society. Nor does it pronounce on inequalities between fellow citizens who all enjoy freedom from domination, and have a secure sense of self-worth and the attitudes required for maintaining a just society. In these cases, by contrast, brute luck egalitarianism has clear distributive implications.

Both ideals of equality seem to me to highlight significant moral concerns, which it is perfectly natural to care about simultaneously. Moreover, as we have understood them, each implies that certain inequalities about which the other remains silent are objectionable. This means that the demands of the two ideals do not necessarily conflict: both ideals may well be fully satisfied by a non-empty subset of feasible social arrangements, and these subsets may show some overlap, so that each ideal rules out some, but not all of the arrangements that the other permits. (Of course, the two ideals may conflict. For example, the only way to achieve brute luck

equality may be to offer all individuals an option set which, through individuals' choices from this set, would lead to a distribution which undermined the conditions of equal citizenship.) It therefore seems to me that one should be both a brute luck and a social and political egalitarian, and that this does not involve endorsing two essentially conflicting ideals.

This conclusion, however, has been disputed by some recent defenders of the social and political ideal of equality. Elisabeth Anderson (1999) and Samuel Scheffler (2004), for example, argue that—so long as this is consistent with equal option sets—the ideal of brute luck equality requires that we leave individuals free to make choices through which they may end up so badly that it is impossible for them to function as free and equal citizens, even when we could make arrangements that would restore them to the minimum condition required at modest cost. To illustrate their complaint, consider the following example. Suppose that equal option sets among a group of well-informed and competent adults could be achieved by providing everyone with the same amount of money and then letting them engage in whatever actions they choose without any state intervention. Some of these individuals might end up indigent and at the mercy of others' aid through imprudent choices, or simply through bad option luck. In this case, Anderson (1999) and Scheffler (2004) argue, the ideal of brute luck equality requires abandoning these individuals to their fate, even if we could, at relatively modest cost, have set up social arrangements in advance which would ensure that such individuals would be guaranteed the conditions required for equal citizenship. (The fact that these arrangements, and any concomitant taxes and transfers, would be set up and announced in advance means that they would not violate anyone's legitimate expectations.)

This criticism is, I believe, based on a misunderstanding of the implications of brute luck egalitarianism. Brute luck equality is consistent with, but does not require, inequalities due to people's choices. It is compatible both with equality of outcome, and large inequalities due to people's choices from their equal option sets. In this example, therefore, brute luck equality would be consistent with an arrangement that set up a social safety net in advance.

Anderson's and Scheffler's criticism is therefore properly directed not against brute luck equality, but against a view of freedom of contract and responsibility that leads us to a particular view of the options people should have, and the consequences individuals face when they choose their options. I discuss several views of responsibility and their relation to brute luck equality in the next chapter.

Chapter 3. Responsibility²⁶

Introduction

In this chapter, I aim to make some headway towards developing a forward-looking view of *substantive responsibility*—the way a person’s claims on others, others’ claims on her, and her level of welfare should depend on the opportunities she has and the choices she makes. I follow Scanlon (1995; 1998) in distinguishing substantive responsibility from two other forms of responsibility. First, *responsibility for one’s judgement-sensitive attitudes* is the sense in which people can be called on to defend their judgement-sensitive attitudes—beliefs, intentions, hopes, desires, and other attitudes like admiration, contempt, etc.—with reasons and to modify them if an appropriate defence cannot be provided (1998, pp. 21-22; 272 and 278). Second, *responsibility as attributability* (which Scanlon also refers to as simply ‘moral responsibility’) is the sense in which it is appropriate to take a person’s judgement-sensitive attitudes or actions as the basis for moral appraisal (1998, p. 248). In what follows, whenever I speak of ‘responsibility’, I will be referring only to substantive responsibility.

The questions I will be studying in this chapter will be of an entirely forward-looking character: I will be focusing exclusively on the question of how we should set

²⁶ I thank Cécile Fabre, Marc Fleurbaey, Michael Otsuka, and Jonathan Wolff for comments on earlier versions of this chapter. I have also benefited from conversations with Tim Scanlon. Earlier versions of this chapter were presented at the Popper Seminar at LSE in November 2004, and at the Workshop on Law and Economics at the World Congress of Legal and Social Philosophy in Granada in May 2005. I am grateful to participants in these sessions for their comments.

up arrangements that will make people's claims, obligations and situation depend on their options and choices.

Substantive responsibility and opportunity-egalitarianism

A view of substantive responsibility is an essential component of an opportunity-egalitarian conception of justice. Brute luck equality demands the absence of inequalities for which individuals cannot be held responsible. It therefore permits, but does not demand, inequalities due to choices for which individuals can be held responsible. Brute luck equality is consistent with equality of outcome; it is also consistent with large differences in welfare that are due to choices for which individuals can be held responsible. We need a principle of substantive responsibility to decide which differences in welfare due to individuals' choices are permissible, or even required.

The issue can also be put as follows. In the previous chapter, we saw that brute luck equality requires providing everyone with equal option sets. Now, we may be faced with several possible ways of providing people with equal option sets. For example, we might be able to provide everyone with either *option set 1*, consisting of two options, one which yields an excellent level of welfare, and one which yields a very low level of welfare, or with *option set 2*, also consisting of two options, both of which yield reasonably good levels of welfare. We would then need a principle of responsibility to choose which of these two option sets to provide everyone with, and thereby determine the content of people's option sets. (Of course, welfare outcomes might not be the only relevant aspect determining the choice between two

option sets. Among other things, the degree of autonomy they allow might play a role as well.)

The role of options and choice in justification

A central question that arises when discussing responsibility is about the role that a person's opportunities to choose and his choices play in the process of justification of arrangements. A person's options and choices can play at least two kinds of justificatory roles. First, a person's options and his choices can be an important determinant of various ways in which his situation is a valuable one for him. (We will look at these ways more closely in section 3.1.) I will call this role of a person's options and choices their 'derivative role', since their role in the justification of arrangements derives from the value that options and choices have for a person.

Second, the very fact that a person has certain options and makes certain choices can play a justificatory role independently of the value that a person achieves by having these options and choices. Consider, for example, someone with adequate capacities of decision-making and decision-implementation, who is well-informed about the options open to her, and who faces an option set with a reasonably good option with a certain outcome which she passes up in favour of an option through which she ends up badly. One might say that in this case, what matters morally is not just the value this person achieves when faced with this option set, but also the value she could have achieved by choosing differently. I will call this role of a person's options and choices in justification their 'fundamental role'.

There are theories of substantive responsibility that give a derivative, but not a fundamental role to a person's opportunities to choose and his choices in the

justification of arrangements. Equality of welfare is an example of a theory of this kind.²⁷ At the fundamental level—the level of justification of arrangements—a welfare egalitarian of this kind is only concerned with the degree of inequality in people’s welfare, and not with how this welfare came about. This means that a person’s options and choices play no fundamental justificatory role. As a consequence, a welfare egalitarian would justify practical arrangements that make people’s claims on others and their obligations to others and the quality of their situation depend on their opportunities and choices by appealing only to the fact that these arrangements brought about the least possible inequality of welfare.

But one might also hold that a person’s options and choices should play a fundamental justificatory role. The precise form this justificatory role can take may be different across different moral theories, of course. But the key idea would be that arrangements which offered people various options under adequate conditions of choice might not only be justified by an appeal to how people end up under these arrangements, but also with reference to the quality of the options they had, but possibly did not choose.

The contrast between theories that do and theories that do not give a fundamental justificatory role to people’s options and choices can be illustrated by the following example, involving the removal of hazardous material from a town. (The example is a modification of an example in Scanlon (1998, chapter 6).) This hazardous material is present in the soil of a particular site through natural causes, and no one could have known about it until its recent discovery. The consequences of leaving it where it is would be very serious—involving, say, the death of a significant number of inhabitants. The material can be removed to a different

²⁷ There are, of course, many other examples of theories of this kind; utilitarianism and prioritarianism among them.

location; once it is deposited there, the health risks will be eliminated. The digging up of the material and its transportation will, however, inevitably release pollutants into the air, which would cause severe, incurable health damage to anyone who is exposed to them by being near the site of excavation or outdoors during its transport. There is no damage to the health of people who stay away from the site and remain indoors during transport.

Given the consequences of leaving the material where it is, the town officials have to remove it. Suppose these officials have to choose between two possible courses of action. Both courses of action, let us suppose, involve significant efforts to limit the amount of pollutants released into the air and thereby decrease the harm to people—other than the workers, who are wearing protective clothing—who are present at the excavation site or outside during transport. But they differ in the following respect. The first policy involves clearly informing everyone of the danger. Suppose that from previous experience in such cases in towns of this size, the members of the town council know that it is very likely that this will mean that almost everyone will stay away from the excavation site, and stay indoors during transport. They also know that it is nearly certain that there will be one member of the population—whose identity cannot be determined in advance—whose impetuous curiosity will be piqued by the warnings, and who will visit the excavation site in spite of the predicted damage to her health. Let us call this person the Impetuous Woman. As a consequence of this policy, and through her informed choice, this woman will come to significant harm.

The second course of action is to divert some of the resources used to ensure that absolutely everyone is informed to building a solid and unclimbable fence around the excavation site. Suppose that from previous experience in such cases in

towns of this size, town officials know that the Impetuous Woman will now not come to harm, as she is barred from visiting the site and has no interest in being outside during the transport. They also know that it is very likely that as a consequence of spending less money on informing everyone, it is nearly certain that one person—whose identity cannot be determined in advance—will not be informed. Unaware of the danger, he will be outside during transport and come to harm. However, if he were warned, he would act prudently and stay indoors during the transport. Call this person the Careful Man.

Now, suppose that the health damage to both the Impetuous Woman, under the first policy, and the Careful Man, under the second policy, is the same, and that their welfare outcomes under each policy are given in table 3.1. Suppose further that everyone else’s situation and options are identical under both policies.

Table 3.1 Two individuals’ welfare under two possible arrangements

Policy	Inform Everyone	Unclimbable Fence
Individuals		
Careful Man	10	6
Impetuous Woman	6	10

Now, in this case, a theory like welfare egalitarianism, which does not give a fundamental justificatory role to people’s options and choices, and which cares only about the distribution of individuals’ welfare, will be indifferent between both policies. (Note that this of course need not mean that, at the level of practical arrangements, there is no substantive responsibility: under both arrangements, people’s outcomes depend on their choices.) By contrast, a theory that does give a

fundamental justificatory role to people's options and choices might discern a morally relevant difference between the two arrangements. For a theory of this kind might say that under the Inform Everyone policy, the Impetuous Woman would have the option of avoiding the bad outcome she would suffer through informed choice. By contrast, under the Unclimbable Fence policy, the Careful Man would not in be a position to avoid the bad outcome he suffered through his informed choice. Given this difference between the conditions in which the two people are placed under the different policies, such a theory might prefer the Inform Everyone policy to the Unclimbable Fence policy. (We will examine various ways in which this preference might be justified in sections 3.1 to 3.3.)

Scanlon on substantive responsibility

The central sections of this chapter focus on two questions. First, what exactly explains the appeal to a person's options and choices in theories that give them a fundamental justificatory role? Second, what are the reasons for favouring a theory that gives a person's options and choices this fundamental role over one that only gives it a derivative role?

My investigation will proceed through a critical examination of Scanlon's (1995; 1998, chapter 6) views on substantive responsibility. In discussing his views, I will diverge from his own presentation of them in one respect. Scanlon is, I believe, concerned with defending a forward-looking view of substantive responsibility and in addressing the question of the justificatory role of a person's options and choices. However, Scanlon's presentation of his central example—which is very close to the aforementioned case of hazardous waste removal—is not ideally conducive to these

aims. For he starts out by assuming that the town council has done all it can reasonably be expected to do to warn and protect people, and that nonetheless, certain people have come to harm, for different reasons: because, moved by impetuous curiosity, they visited the excavation site, or because they were uninformed in spite of the council's efforts, or because they simply forgot the warnings they were given. He then asks how we should defend the arrangements the council has made to the people who have come to harm, and what role their options and choices play in this defence. This way of presenting the example is not entirely satisfactory, since by focusing on after the fact analysis, it draws attention away from the central issue, namely how the town council should decide when faced with different possible courses of action, and what role people's options and choices should play in making this decision. I have therefore re-cast Scanlon's example as a decision problem for a town council which is trying to decide between different arrangements under which different people, through different processes, will come to harm.

In the introduction to *What We Owe to Each Other*, Scanlon claims that

“the force of a person's reasons for rejecting a principle that would require him to bear a certain burden can be reduced by the fact that this burden is one he could have avoided by choosing appropriately” (1998, p. 9)

And again, in introducing his discussion of substantive responsibility, he writes:

“the force of a person's objection to a principle imposing a burden on her, or permitting others to act in a way that would impose such a burden, can be diminished by the fact that she could avoid that burden by choosing appropriately” (1998, p. 249).

From the outset, Scanlon therefore appears to accept that a person's options and choices should play a fundamental role in justifying moral principles for the regulation of our behaviour. For this reason, his discussion focuses not on whether a person's options and choices should play this role, but on the precise nature of this role. He advances two contrasting theories of this justificatory role of a person's options and choices, which he calls the Value of Choice view and the Forfeiture view, and argues in favour of the former. Very roughly, Scanlon draws the contrast between the two as follows. On the Value of Choice view, what matters is the general quality of a person's opportunities to choose. By contrast, the core idea of the Forfeiture view is that someone placed in adequate conditions of choice (with the requisite information and decision-making and decision-implementing capacities, etc.) who has an accessible, secure and prudent option, and who makes a conscious choice to pass up this option, cannot complain if she ends up badly as a result of her choice: *volenti non fit iniuria*.

Outline of the chapter

My discussion will proceed as follows. In section 3.1, I discuss the Value of Choice view. I argue that Scanlon offers no good reason for evaluating a person's option set by the *general* goodness or badness of that option set, rather than how valuable it is for that person, given her personal characteristics. I also discuss an alternative method of evaluating a person's option set, which I call the Realised Value of Option Set view, which equates the value of an option set with the value a person actually realises by having that option set. I show that this method gives a person's choices and options a derivative, but no fundamental justificatory role.

In section 3.2, I discuss the Forfeiture view. I argue that though it gives a person's choices and options a fundamental justificatory role, it is unattractive because it pays insufficient attention to the value of removing disadvantageous options from a person's option set.

In section 3.3, I present an outline of a view of responsibility that gives a person's options and choices a fundamental justificatory role whilst avoiding the Forfeiture view's problems. I call this perspective the Potential Value of Option Set view. Very roughly, this view proposes that when considering a person's claims, we consider the value of her option set. The value of her option set is a function of the various values which she can achieve by choosing from it. It is also a function of the ease or difficulty with which she can achieve these values by making different choices from her option set.

Since I argue that we should reject the Value of Choice view and the Forfeiture view, we are left with the question whether we should adopt the Realised Value of Option Set view, and give a person's options and choices a derivative role only, or adopt the Potential Value of Option set view, and give them a fundamental justificatory role as well. In section 3.4, I tentatively suggest that an important consideration in this regard is the truth or falsity of determinism: the thesis that "the prevailing laws of nature are such that there do not exist any two possible worlds which are exactly alike up to some time, which differ thereafter, and in which those laws are never violated" (Lewis 1973, p. 559). For it seems that giving a person's options and choices a derivative role only is compatible with the truth of determinism, whereas giving it a fundamental role in justification may not be.

In section 3.5, I consider the implications of the forgoing for an opportunity-egalitarian conception of justice.

3.1 The Value of Choice account

The ways in which choice can be valuable

The Value of Choice account, in Scanlon's words, focuses on "the positive reasons that people have for wanting opportunities to make choices that will affect what happens to them, what they owe to others, and what others owe to them" (1998, p. 251). Scanlon distinguishes three generic reasons for wanting what happens to us to depend on our choices. The first is what he calls the *instrumental* value of choice: the value choice has in securing states of affairs that we seek. This value of choice is conditional on the degree to which for a given object of choice, a person's capacities and conditions of choice will help him achieve his ends. It is also relative: it depends on the usefulness of his being given a choice as compared to other means of achieving his ends.

The second is the *representative* value of choice. This is the value we put on seeing features of ourselves manifested in our actions and their results. Examples are gifts, where the significance of the gift is determined by having chosen it oneself, and creative work, where part of the point of the work is that it reflects its author's attitudes and abilities.

The third is the *symbolic* value of choice. We may want outcomes to depend on our choices not merely because this will be a more efficient way of achieving our ends, or because we want our choices to reflect our values, thoughts and capacities, but also because not making such choices ourselves would be taken as an indicator

that we are not competent or do not have the standing of a normal adult member of society.

Scanlon (1998, p. 253) does not take this list to be exhaustive, and I think it is important to add at least two other ways in which choice can be valuable. Choice can be valuable because of its contribution to a person's *autonomy*. Having a range of different choices and responsibility for the concomitant outcomes is valuable because it is one of the necessary conditions for an autonomous life. As Raz puts it:

“The ruling idea behind the ideal of personal autonomy is that people should [be (part) authors of] their own lives. (...) The ideal of personal autonomy is the vision of people controlling, to some degree, their own destiny, fashioning it through successive decisions through their lives (...) It is opposed to a life of coerced [or manipulated] choices. It contrasts with a life with no choices [because of lack of options] or of drifting through life without ever exercising one's capacity to choose” (Raz 1986, pp. 369 and 371).

Autonomy, then, places certain demands on people's range of options. Firstly, they must be sufficiently diverse. An individual faced with very similar options will have no incentive to consider different aims and pursuits; consequently, he will also be unlikely to develop his capacities of evaluation and choice. For the same reason, the range of options must involve several reasonable options. If all but one of a person's options involve horrible consequences, then again, he will have no reason to seriously consider different alternatives and no real ability to shape his life according to his judgements (Raz 1986, p. 373).

These four forms of value represent ways in which facing a certain option set and making a particular choice can be valuable for the person with this option set. But of course, giving people certain options with certain payoffs can have *value for others*, too.

It may be to the advantage of someone's employers, for example, to have her pay depend on her choices, like the amount of hours she works, and the dedication she brings to the job, independently of whether this would be to her own advantage. Similarly, removing certain choices may be costly for some. A case in point is the previously mentioned policy of placing an unclimbable fence around the site of the excavation of hazardous material. This would remove the choice of visiting the excavation site, at the cost of resources that could be used for other purposes, like informing everyone of the danger of being outside during the transport of the hazardous material. In this case, getting very close to the excavation site could not be seen as a valuable activity for anyone, no matter how curious they were. Therefore, not placing the fence because of its cost to others, and thereby giving people the option of getting too close to the site would be a form of assigning substantive responsibility not because of the value of this choice to the chooser, but because of the benefits that giving people this option would bring others.

Of course, identifying these different ways in which a person's having certain options and making certain choices can be valuable for her or for others is not enough. We also have to make intrapersonal tradeoffs between the first four values: giving someone a choice in one set of circumstances may have positive symbolic value for her, for instance, but also have negative instrumental value for her.

Consider again the case of hazardous waste site. Suppose that building an unclimbable fence were not an option, but that instead we could spend the same resources on profiling certain types of people whom we know would be likely to be foolish enough to get too close to the site. Suppose we could prevent their coming to harm by requiring that they be accompanied by others on their daily outings during this period. (This is, of course, what is often done for children and teenagers, who

are sometimes placed under curfew and have to be accompanied by responsible adults during activities in which they might come to harm through their tendency to risky behaviour.) Compared to the Inform Everyone policy, which involved no such restrictions on a particular group of people, this Profiling Policy might have negative symbolic value for these people, since they would not be singled out for their likely imprudent behaviour. But if they would indeed come to harm under the alternative Inform Everyone policy, then the Profiling Policy would have positive instrumental value for them.

We also have to consider tradeoffs between how different people are affected by particular arrangements. In our comparison between the Inform Everyone and Unclimbable Fence policies, for example, we have to trade off the welfare of the Impetuous Woman against that of the Careful Man.

Here, I want to leave aside questions about how to make the tradeoffs between the different ways in which having options and making choices can be valuable for a person, and questions about how, in general, we should weigh different individuals' claims. Instead, I want to focus on just one question: on the Value of Choice view, what is the nature of the justificatory role of a person's options and choices?

The role of options and choices in justification in the Value of Choice view

As we saw, Scanlon appears to give a person's options and choices this fundamental justificatory role. Scanlon argues that the Value of Choice account can do so by tying justification to the value of the option set that a person is presented with:

“On the Value of Choice account what matters is the value of the opportunity to choose that the person is presented with. If a person has been

placed in a sufficiently good position, this can make it the case that he or she has no valid complaint about what results (...)” (1998, p. 258).

How does Scanlon determine the value of a person’s option set? One way of determining the value of a person’s option set is simply to equate it with the value it yields for her—the value she will achieve both by having the option set and by choosing the option she does from his option set. Let us call this the Realised Value of Option Set view. Let us suppose for simplicity that in our example, choice has only instrumental value, and that we judge an option’s instrumental value by the welfare it yields, so that only the welfare outcomes of these choices are relevant to the evaluation of the quality of her option set (Scanlon (1998, p. 257) also makes this assumption in this case). This means that we set the value of individuals’ option sets to the value of their welfare outcomes, so that the information relevant to evaluating the two policies is given by table 3.1.²⁸

Now, on the Realised Value of Choice view, a person’s options and choices obviously only have a derivative role in justification. Given that Scanlon is attempting to provide an explanation for the fact that, as he puts it, “the force of a person’s objection to a principle imposing a burden on her (...) can be diminished by the fact that she could avoid that burden by choosing appropriately” (1998, p. 249), he must not mean to regard the value of a person’s option as equivalent to the value the person achieves when she has the option set.

²⁸ Obviously, on this approach, we might not always take the value of a person’s option set to be equivalent to the value of the option she chooses. For being given an option set, rather than just obtaining the welfare associated with the chosen option, may itself have value for the person; it may, for example, enable her to be more autonomous.

Indeed, Scanlon does not tie the evaluation of a person's option set to how good it is for each individual. He argues that we can distinguish between the value we put on a person's option set and the value it has for that person by attending to the distinction between "the generic reasons on which the justifiability of a moral principle must rest and the reasons that a specific individual may have, given all the facts about his or her situation" (1998, p. 263). These "generic reasons" are reasons "that we can see that people have in virtue of certain general characteristics" (1998, p. 205). Applied to the hazardous waste removal case, this means that since it is generally in people's interest to be warned of the dangers of visiting the excavation site, receiving this warning will count as improving everyone's circumstances of choice, even if they are the idiosyncratic type of person whose tendency to commit imprudent acts is thereby increased. As Scanlon writes of the case of the Impetuous Woman under the Inform Everyone policy:

"The reason why it is important that this woman was informed of the danger, and thus given the chance of avoiding it, is not that this is necessarily advantageous to her but rather that it is something that people in general have reason to value and hence to demand that an acceptable principle insist on" (1998, p. 263).

Scanlon argues that this evaluation of a person's option set in terms of generic reasons allows us to register an important moral difference between the situation of the Impetuous Woman under the Inform Everyone policy and the situation of the Careful Man under the Unclimbable Fence policy. Of the case of the Careful Man when he is uninformed, he writes that his option set is less good than that of the Impetuous Woman when she is informed but is not effectively prevented from

visiting the excavation site, because he does not receive the benefit of being informed, while she does:

“because we did not succeed in making him aware of the danger, we did not make what happened to him depend on his response to this information. Given that this dependence is something we all would reasonably want to have under the circumstances, we did not succeed in making this person as well off as he would reasonably want to be. The [Impetuous Woman], however, did have the benefit of being informed, even though this turned out to be worth less as protection than it would have been to most other people” (1998, p. 259).

Now, as mentioned in this chapter’s introduction, Scanlon does not explicitly consider the choice between two policies, under one of which the Impetuous Woman comes to harm, and under another of which the Careful Man suffers the same harm. But given his comments on the quality of their respective circumstances of choice, we appear to be able to draw the conclusion that we should choose the Inform Everyone policy. For the value of the Careful Man’s option set under the Inform Everyone Policy and the Impetuous Woman’s option set under the Unclimbable Fence policy would appear to be equal (both people are well-informed and well-placed to achieve the best outcome, that is, staying indoors during excavation and transport). But, according to Scanlon, the value of the option set of the Impetuous Woman under the Inform Everyone Policy exceeds that of the Careful Man under the Unclimbable Fence policy, since she is informed and he is not, and being informed amounts to a benefit. (Below, I will question Scanlon’s view that being informed amounts to a benefit to the Impetuous Woman. At this point, I am simply examining the conclusions we can draw from Scanlon’s discussion.) This

means that, if we assess individuals' situations in terms of the value of their option sets determined in this way, the Inform Everyone policy is unambiguously better, since the most valuable option set under this policy is as valuable as the most valuable option set under the Unclimbable Fence policy, and the least-valuable option set is more valuable. This evaluation of individuals' option sets is represented in table 3.2. The discussion of this paragraph has established that $x > y$ and $z > y$.

Table 3.2 A Scanlonian evaluation of individuals' option sets under two possible policies

Policy	Inform Everyone	Unclimbable Fence
Individuals		
Careful Man	x	y
Impetuous Woman	z	x

Criticism

Scanlon's method of evaluating a person's option set, and his ability to distinguish between the claims of the Careful Man when he is uninformed and the Impetuous Woman when she is able to visit the excavation site depend crucially on his appeal to what are generally good circumstances of choice as opposed to the value that these circumstances of choice have for each specific individual. We also saw that Scanlon justifies his use of generally good circumstances of choice by appealing to the fact that moral evaluation has to make use of what he calls "generic reasons". I will now argue that this appeal to generic reasons cannot justify his method of evaluation.

Scanlon has three reasons for appealing to generic reasons in evaluating a person's situation. The first reason is that taking into account specific variations in people's needs and circumstances would be more demanding than just paying attention to general characteristics; it would lead to greater uncertainty about whether everyone's claims had been met and require everyone to gather more information in order to know what a principle gives to and requires of them (1998, p. 205). Since this uncertainty and information gathering are costly, we are permitted to refer to generic reasons in the justification of our actions and principles of action.

In my design of the hazardous waste removal case, I supposed that by ordinary methods (research into what happened in comparable situations elsewhere), town officials would know that it was virtually certain that under the Inform Everyone policy, there would be a person like the Impetuous Woman who would come to harm in this way, and that under the Unclimbable Fence policy, there would be a person like the Uninformed Man, though I also assumed that they did not know who these people would be (or, indeed, whether they would be male or female). Given this assumption, the argument from the cost of information cannot explain why we should consider the Impetuous Woman as having received a benefit by being informed in the Inform Everyone case. For there is no additional cost involved in evaluating her option set by the value it yields for her, which is, as we know, just that it leads to health damage.

Scanlon offers a second reason for attending to generic reasons in an early paper, 'Preference and Urgency' (1975). There, he argues that not all of a person's preferences give rise to moral claims on us. The case of the Impetuous Woman might offer an example of that kind. For no matter how strong her desire to see the excavation site, one might argue that the satisfaction of this desire does not give rise

to a claim on the council for its satisfaction that needs to be balanced against the harm that will befall her when she visits the excavation site. This view implies that we should assess the quality of a person's options in terms of certain values, which might diverge from the person's own value judgements. In evaluating the quality of the Impetuous Woman's options in the Inform Everyone case, for example, we would not be required to consider the fact that she satisfies her curiosity if she visits the excavation site as a benefit to her, which would add slightly—or even significantly—to her level of welfare if she were to choose this option.

Now, this view offers us no reason to regard the Impetuous Woman as having been placed in good circumstances of choice in the Inform Everyone case. For, by Scanlon's own hypothesis, being given an informed choice in this case has only instrumental value, and in this case it has *negative* instrumental value for the woman, since, as Scanlon supposes, "the warning only aroused her impetuous curiosity, and she would have been better off if she had never been told at all" (1998, p. 261). So there is no generic reason to regard being informed as a benefit to her. Moreover, there is a generic reason to regard her circumstances of choice as being bad. For she will suffer damage to her health, and this among the ways in which a person is affected that Scanlon, in 'Preference and Urgency', regards as giving rise to legitimate claims.

On the basis of our discussion in chapter 1, we can also see that there is a third reason to sometimes disregard the particular reasons a person has for valuing certain things and taking certain actions. For I argued there that there is no plausible way of evaluating a person's welfare that will always follow that person's judgements of her own good. I also argued that we should follow Scanlon's (1991) proposal and adopt a substantive measure of welfare based on a list of goods and conditions that are

recognised as valuable by people with different values. This means that when we evaluate a person's welfare, we must sometimes make judgements that will conflict with her own, and therefore must sometimes disregard her particular reasons for valuing or wanting something.

Once again, accepting this reason for disregarding the specific reasons an individual may have does not offer us any grounds for regarding the Impetuous Woman's option set in the Inform Everyone case as a good one. For, as mentioned in section 1.4, a person's health is a good candidate for figuring in this list of goods and conditions that are constituents of a public measure of welfare. Thus, like the second reason for appealing to generic reasons, this reason justifies using a public standard of value that may differ from her own standard to assess the value she can or does obtain through her option set. But it does not justify assigning to her option set a value greater than the one she actually obtains from it; that, it seems, can only be justified by an appeal to her power to choose differently than she did.

In sum, we should reject Scanlon's reasons for not evaluating a person's option set by the value it yields for her. What are the alternatives to Scanlon's method of evaluation? We have already encountered one such alternative: the Realised Value of Option Set view, which evaluates a person's option set by the value it yields for her. As we have seen, this would give a person's options and choices a derivative, but not a fundamental role in the justification of arrangements. It therefore cannot explain what Scanlon sets out to explain, namely, how "the force of a person's objection to a principle imposing a burden on her (...) can be diminished by the fact that she could avoid that burden by choosing appropriately" (1998, p. 249). We will examine a different attempt to explain this role of a person's choices and options in the next section.

3.2 The Forfeiture view

Scanlon characterises the central idea of the Forfeiture view as follows:

“a person who could have chosen to avoid a certain outcome, but who knowingly passed up this choice, cannot complain of the result: *volenti non fit iniuria*” (1998, p. 259).

Concerning the case of the Impetuous Woman under the Inform Everyone policy, Scanlon writes that a proponent of the Forfeiture view would reason as follows:

“Since she had been warned of the danger, and chose to go to the site anyway, we are inclined to say that she is (substantively) responsible for her own injury; and it is this fact, rather than the amount that has been done to protect her or the cost to others of doing more, that makes it the case that she cannot blame anyone for what happened. By choosing, in the face of warnings, to go to the (...) site, she laid down her right to complain of the harm she suffered as a result” (1998, p. 258).

The core idea of the Forfeiture view, then, is that someone placed in adequate conditions of choice (with the requisite information and adequate decision-making and decision-implementing capacities, etc.) who has an adequately accessible prudent option which it would be reasonable for her to take, and who makes a conscious choice to pass up this option, cannot complain if she ends up badly as a result of her choice. Scanlon says little more about the Forfeiture view, so that it is not wholly clear which decision principle we should take to represent this view of responsibility. But the quoted passage suggests the following: we should base each person’s moral

claims on the quality of her option set, which is determined as follows. If a person is adequately informed, and has adequate decision-making and decision-implementing capacities, and her option set contains an adequately accessible prudent option which it would be reasonable for her to take, then take the value of her option set to be equal to the value she would achieve by choosing this prudent option. When the conditions of information and rational capacities are not met, evaluate the person's option set by the value she achieves through it. This would lead to the evaluation of the Impetuous Woman's and the Careful Man's option sets listed in table 3.3.

Table 3.3 The Forfeiture view's evaluation of individuals' option sets under two possible policies

Policy	Inform Everyone	Unclimbable Fence
Individuals		
Careful Man	$x = 10$	$y = 6$
Impetuous Woman	$z = 10$	$x = 10$

This evaluative procedure would obviously make Inform Everyone preferable to Unclimbable Fence. Thus, by basing individuals' claims on the quality of their option set, and evaluating each person's option set not by the value she achieves, but by the value she could achieve if she made a reasonable and prudent choice, the Forfeiture view clearly gives a person's options and choices a fundamental justificatory role.

Scanlon's criticism of the Forfeiture view

Scanlon offers two points of criticism of the Forfeiture view. The first is that the Forfeiture view places undue weight on the special legitimating force of voluntary action (1998, p. 260). What is important, Scanlon claims, is the opportunity to choose that a person has, rather than his conscious decision to pass up specific alternatives. Scanlon argues that we should appeal to the former rather than the latter because there are cases in which it seems right to offer the options a person has as part of the justification of an arrangement under which he ends up badly through an action of his own, even if he never consciously considers all relevant aspects of this action. We can illustrate cases of this kind as follows. (This is a variation on an example offered by Scanlon (1998, p. 259).) Imagine that in our hazardous waste removal example, under both the Inform Everyone and Unclimbable Fence policies, the town officials can confidently predict that there will be a person—whose identity, again, they do not know—who, though informed of the risk of contamination, will simply forget about it. (Call this person the Forgetful Man.) As a result, the Forgetful Man will be outside during the transport of the hazardous material and suffer damage to his health.

Imagine further that there is a third possible policy, in which we use the resources that otherwise would be used to inform everyone or build an unclimbable fence to issue the warning in a way that is particularly easy to remember. Under this policy, the Forgetful Man would remember to stay inside during the transport of the material, though the Careful Man and Impetuous Woman would each come to harm. The results of the three policies are listed in table 3.4.

Table 3.4 Three individuals' welfare under three possible policies

Policy	Inform Everyone	Unclimbable Fence	Easy to Remember
Careful Man	10	6	6
Impetuous Woman	6	10	6
Forgetful Man	6	6	10

Scanlon argues that the case of the Forgetful Man illustrates that it is not active choice, but the quality of the circumstances in which a person was placed, that is of moral significance. For, he argues, the Forgetful Man does not consciously decide to take a dangerous action; nevertheless, if he is warned and has normal cognitive capacities, then, Scanlon believes, his claims should be treated like those of the Impetuous Woman, who makes a conscious decision to take the dangerous action (1998, p. 259). (This would mean, for example, that we should be indifferent between Unclimbable Fence and Easy to Remember.)

I see two problems with Scanlon's point against the Forfeiture view. First, it is not clear whether the case of the Forgetful Man reveals that informed choice is not the crucial moral fact. If we assume the Forgetful Man has adequate decision-making and decision-implementing abilities, then though he does not make a conscious decision to engage in dangerous activity, and while we may assume that he does not 'choose to forget' the relevant warnings, he does make a conscious decision about how to deal with the warnings. He could, for example, write the date and time of the transport on his calendar as soon as he hears them; or assemble other reminders. But, we may assume, he consciously chooses not to do so. It is this choice, one might

argue, that makes it right to refer to the opportunity he has to avoid coming to harm when deciding which policy to pursue.

Secondly, Scanlon's definition of the Forfeiture view specifies that informed choice under certain circumstances is a sufficient, and not a necessary condition for not being able to complain of a result. The case of the Forgetful Man therefore does not constitute a counterexample to the Forfeiture view. If Scanlon is right that no conscious choice is involved in the case of the Forgetful Man, then all this proves is that under certain conditions, both informed, voluntary choice and negligence can give grounds for not being able to complain of a result of one's actions. We should then redefine the Forfeiture view as follows: someone placed in adequate conditions of choice (with the requisite information and adequate decision-making and decision-implementing capacities, etc.) who has an adequately accessible prudent option which it would be reasonable for him to take, and who passes up this option either through negligence or conscious choice, cannot complain if he ends up badly as a result.

Redefined in this way, the Forfeiture view still gives a person's options and choices a fundamental justificatory role by evaluating each person's option set not by what he achieves, but by what he could achieve if he chose reasonably prudently.

Scanlon's second objection to the Forfeiture view runs as follows. He points out that the Forfeiture view must stress the fact that the Impetuous Woman could have done otherwise than she did. But, he argues, identifying this fact as the crucial one leads to implausible results in other cases, since, he argues,

“there are many conditions that undermine the legitimating force of choice despite the fact that a person choosing under such conditions still ‘could have done otherwise’ in any sense that would apply in this case” (1998, p. 262).

He gives the following example of such a case:

“It would, for example, be reasonable to reject a principle according to which a long-term contract is binding even when entered into by a fourteen-year-old without adult guidance. What is special about the case of fourteen-year-olds is not that they *cannot* choose wisely (after all, many of them do), but rather that they are so likely not to do so” (1998, p. 262; emphasis in original).

The challenge Scanlon puts to the Forfeiture view is essentially to explain what makes it the case that informed, voluntary choice does not play a justificatory role in the case of fourteen-year-olds when it does play this role in the case of the Impetuous Woman, while not referring to the fact that fourteen-year-olds are unlikely to choose wisely—for that is true too of the Impetuous Woman.

A defender of the Forfeiture view could respond that the Impetuous Woman, like other normal adults, is assumed to have a certain amount of knowledge, experience, and certain cognitive capacities, like the ability to reflect coolly on what she will do and fully appreciate the potential consequences, and a degree of will-power that fourteen-year-olds typically lack. This knowledge and these capacities are, like the availability of a reasonable and prudent option, necessary conditions for choice to have the moral force the Forfeiture view accords to it. Though, as the Forfeiture view theorist will readily acknowledge, these conditions usually make it likely that a person will choose well, their import is not reducible to their contribution to a person’s choosing well, or to any other value that choice has. They simply make it reasonable for us to ask of people that they ‘look out for themselves’; and if they do not, that need not be of concern to us.

In sum, Scanlon’s second objection lacks force because a defender of the Forfeiture view can happily agree with Scanlon that “there are many conditions that

undermine the legitimating force of choice despite the fact that a person choosing under such conditions still ‘could have done otherwise’ ”.

Further criticism of the Forfeiture view

A more forceful objection to the Forfeiture view, I believe, begins with the observation that on this view, so long as the people involved are adequately informed and have adequate decision-making and decision-implementing capacities, and have adequate access to a prudent option which it would be reasonable for them to choose, we are free to disregard the quality of the other options in their option set. For this view of responsibility takes the slogan *volenti non fit iniuria* literally: it completely disregards the value of the options in a person’s option set other than the prudent one. Applied consistently, this view of responsibility would have very unappealing consequences. It would not allow anyone to object to a bad outcome that they could have avoided through reasonable and prudent action, no matter how easy or cheap it would have been for others to prevent this bad outcome from occurring. In the initial specification of the hazardous waste removal example, I supposed that under both policies, significant efforts are made to lessen the amount of pollutants released into the air, so that those exposed by visiting the site or by being outside during transport would be less severely harmed. But it seems that so long as the harm of exposure is reasonably avoidable by everyone, the Forfeiture view, as we have spelled it out, would not require these efforts. Even if, for example, the cost of reducing the harm of exposure was in fact very small in comparison with the harm prevented (say, minimal extra expenditure to reduce the harm to anyone exposed from death to emphysema), the Forfeiture view would not require reducing

the harm, so long as exposure was due only to people's informed, voluntary choices, and they had adequate decision-making and decision-implementing abilities.

As formulated, the Forfeiture view would even permit giving people option sets containing options that it is in no one's interest to have. Imagine a population in which everyone has equal option sets containing at least one good option which it is reasonable to choose. Suppose also that the outcome associated with the worst option these sets contain is very bad, but that we could costlessly remove this option from each person's option set. Suppose also that there is no sense in which being granted the opportunity to choose this option is of value to anyone. Suppose that we know that some people, in spite of being well informed and generally having adequate choice-making abilities, may nonetheless choose the worst option in their set. In this case, the Forfeiture view, which only pays attention to the value of the prudent option, would nonetheless permit not removing this option. For if we take the principle *volenti non fit iniuria* literally, people would have no complaint if they were given this extended option set and chose their worst option.

This view of responsibility is unappealing because it seems perfectly reasonable to demand some expenditure of resources, even at the cost of diminishing the value of people's prudent options, to eliminate tempting but bad options from people's option sets, or to make the outcomes associated with choosing these options less bad.²⁹ To some extent, this may well be to everyone's advantage. For each person knows that, on occasion, he may choose unwisely, and it is therefore generally in his interest to be protected against (the consequences of) making bad choices (cf. Scanlon 1998, p. 263). But even when it is not to everyone's advantage, some such expenditure is warranted. For we are generally not permitted to let others come to

²⁹ The Lord's Prayer, after all, runs "Lead us not into temptation, but deliver us from evil."

great harm through their choices when this harm is easily preventable at comparatively little cost.

3.3 The Potential Value of Option Set view

So far, we have reviewed three ways of evaluating a person's option set: the Value of Choice view, the Forfeiture view, and the Realised Option Set view. We have rejected the first two, and established that the third gives a person's options and choices a derivative role in justification only. There is, I propose, a plausible fourth way of evaluating a person's option set which gives a fundamental justificatory role to a person's options and choices.

Consider again the hazardous waste removal case. Under the Inform Everyone policy, both the Impetuous Woman and the Careful Man know that they have an option set with two options, and they also know the value of these options: 'do not visit the excavation and stay indoors during transport', with a welfare outcome of 10, and 'visit the excavation site or go outdoors during transport', with a welfare outcome of 6. Under the Unclimbable Fence policy, the Impetuous Woman knows she has an option set with two options, and knows the value of these options: 'do not visit the excavation and stay indoors during transport' has a welfare outcome of 10, and 'go outdoors during transport' has a welfare outcome of 6.

Now, let us assume that the option sets under consideration have only instrumental value. Since these option sets contain an option that is strictly better than the least valuable option, the least we can say about the value of these option sets is that they are worth more than the least valuable option. This means that the

value of the three aforementioned option sets is larger than 6. It also seems reasonable to say that since they contain an option that is strictly less valuable than the best option, the value of these option sets is less than the value of the best option, so that the value of the three aforementioned option sets is smaller than 10.

Furthermore, it seems reasonable to hold that the ‘ease’ or ‘difficulty’ with which a person can make each choice in her option set is also relevant to its value. In our example, we know that under the Inform Everyone policy, the Impetuous Woman will be tempted to disregard the danger to her health and visit the excavation site. By contrast, under the Unclimbable Fence policy, she will experience no such temptation to act imprudently, so that taking the prudent action will be easy. It therefore seems appropriate to say that the option set she faces under the Unclimbable Fence policy is more valuable for her than the option set she faces under the Inform Everyone policy. As for the Careful Man, assuming his character is such that he will experience no temptation to act imprudently when he is fully informed of the danger, and since the number and value of the options in their option sets are the same, it seems reasonable to say that his option set under the Inform Everyone policy is of equivalent value to the option set of the Impetuous Woman under the Unclimbable Fence policy. Finally, under the Unclimbable Fence policy the Careful Man has, through no choice or fault of his own, no knowledge of the consequences of his options. We should therefore set the value of his option set to the outcome he achieves through his uninformed choice, which is 6. This assignment of value to individual’s option sets is given in table 3.5, where our conclusions imply $10 > x > z > y = 6$.

Table 3.5 The value of individuals' option sets under two possible arrangements under the second approach to evaluating option sets

Policy	Inform Everyone	Unclimbable Fence
Individuals		
Careful Man	x	$y = 6$
Impetuous Woman	z	x

Now, if we evaluate each policy in terms of the value of a person's option set given in table 3.5, we should prefer the Inform Everyone policy to the Unclimbable Fence policy. For the distribution of the value of people's option sets under Inform Everyone is unambiguously better than the distribution of the value of people's option sets under Unclimbable Fence. For the least valuable option set under Inform Everyone is more valuable, and the most valuable option set is equally valuable.

Let us outline with a bit more precision and generality the five principles for evaluating option sets we have just used.

If someone is adequately informed, and has adequate decision-making and decision-implementing capacities, and has an option set of two or more options, and the option set has only instrumental value, then:

- (i) an option set, which contains at least one option which is strictly better than the least valuable option, is more valuable than the least valuable option;
- (ii) an option set which contains at least one option that is strictly less valuable than the best option is less valuable than the best option;
- (iii) two option sets are of equal value if (a) these option sets contain an identical number of options n , (b) the 1st best, 2nd best, ..., and n th best

options yield equal value, and (c) the ‘ease’ or ‘difficulty’ with which the 1st best option in the first option set can be chosen is equal to the ‘ease’ or ‘difficulty’ with which the 1st best option in the first option set can be chosen, and so on for all options;

and

(iv) a first option set is more valuable than a second option set when (a) these option sets contain an identical number of options n , (b) the 1st best, 2nd best, ..., and n th best options yield equal value, and (c) it is easier to avoid the worse options and to choose the better options in the first option set than it is in the second.

On the other hand,

(v) If a person has only one option or lacks adequate information, or lacks adequate decision-making or decision-implementing capacities, then his option set is evaluated at the value he achieves through it.

As the discussion of our example shows, these principles are jointly sufficient for giving a person’s options and choices under adequate conditions of choice a fundamental justificatory role. (Indeed, principles (i) and (v) together are sufficient for this purpose.³⁰) Moreover, if we accept the idea of evaluating a person’s option set not just by the value he achieves through it, but also by what he could achieve through it, then they seem to be plausible principles of evaluation. However, they are

³⁰ To see this, consider how, using principles (i) and (v), we would evaluate the option set of the Impetuous Woman under a different policy, under which she was not informed of the danger, and, as a consequence was outside during the transport. From the perspective of the value of the Impetuous Woman’s option set, this policy would be worse than the Inform Everyone policy, even though her level of welfare under both policies would be equal.

not yet sufficient to block the objections to the Forfeiture view. For example, they do not imply in the example reviewed in section 3.2, in which we could costlessly remove a bad option from everyone's option set, that removing this option will increase the value of everyone's option set. But these principles are perfectly compatible with further principles of evaluation which could ensure the appropriate judgements in such cases. Such principles should require us to give weights to the value of each option in a person's option set, as well as to the ease or difficulty with which each option can be chosen, in such a manner that protection against choosing badly is considered a valuable characteristic in an option set.

I will not discuss such principles here.³¹ Instead, I hope to have provided a basic sketch of and a route to further development of what I will call the Potential Value of Option Set view. When a person has more than one option, is adequately informed and has adequate decision-making and decision-implementing abilities, this view takes the value of a person's option set to be a function of both the various values which she can achieve by choosing from her option set and of the ease or difficulty with which she can achieve these values by making different choices from her option set. Contrary to the Value of Choice view, it takes into account a person's specific characteristics by accounting for the ease or difficulty with which she can make particular choices. Contrary to the Forfeiture view, it takes the value of a person's option set to be determined by the value of more than just the value a person achieves if she chooses the reasonable and prudent option from this set. Contrary to the Realised Value of Option Set view, it does not always equate the

³¹ For some introductory discussion and references see Sen (2002, especially chapters 20-22). See also Vallentyne (2002), and Ooghe, Schokkaert and Van de gaer (forthcoming). I discuss and criticise one well-known method of assigning value to option sets in chapter 4.

value of an option set with the value a person realises through her choice from that option set. It therefore provides a fundamental justificatory role for a person's options and choices, without disregarding the value of protection against choosing badly.

3.4 The Realised Value of Option Set view, the Potential Value of Option Set view, and determinism

Of the four views of substantive responsibility we have examined, the Realised Value of Option Set view and the Potential Value of Option Set view remain as plausible candidates. The key difference between the two is that in assessing a person's moral claims, only the Potential Value of Option Set view makes reference to what he can achieve through choosing appropriately. This means that if a person would end up badly through his voluntary choice or negligence under a certain arrangement, but he could have been better off if he had acted differently, he may not have the same grounds for complaint against this arrangement as someone who ends up equally badly through no choice or fault of his own. It therefore allows us to register a moral difference between the claims on behalf of the Impetuous Woman under the Inform Everyone Policy, and the Careful Man under the Unclimbable Fence policy. It also allows us to register a difference between the moral claims on behalf of the Forgetful Man under the Inform Everyone policy and the Careful Man under the Easy to Remember policy. In these comparisons, it allows us to say that we can advance a less forceful claim on behalf of the Impetuous Woman and the Forgetful Man than on behalf of the Careful Man, because the former would have an adequate

opportunity to avoid ending up badly, whereas the Careful Man would not. In this respect, it seems to match moral intuitions about the fundamental justificatory role of a person's options and choices which Scanlon takes for granted. These intuitions speak in favour of the Potential Value of Option Set view, since the Realised Value of Option Set view cannot accommodate them.

It is important to note that these intuitions cannot be undermined by drawing attention to equally strong intuitions, which are reflected in everyday arrangements, that we should arrange things collectively and individually to protect individuals like the Impetuous Woman and the Forgetful Man. For, as we have seen, the Potential Value of Option Set view can accommodate the judgement that such protection is morally required. The Potential Value of Option Set view only holds that preventing harm that would befall people through their own choice, in a situation in which they would have the opportunity to avoid this harm, may be less of a moral priority than preventing equal harm to people who would not have this opportunity to avoid it. It in no way implies that preventing harm that would befall people through their own choice, in a situation in which they would have the opportunity to avoid this harm, may be of no moral importance; indeed, I have argued that it should accord the prevention of such harm significant weight.

I believe, however, that the intuitive attractiveness of the Potential Value of Option Set view might be undermined by the thought that determinism might be true. On this point, I take my cue from Scanlon's remarks at the end of his discussion of substantive responsibility, when he characterises the sentiment he believes should underpin a plausible view of substantive responsibility as follows:

“[W]hen we follow a policy that leads to some people's being injured because they have ignored the warnings they were given, we may be correct in feeling

that what we do is justified. But we must also recognise that what separates us from such people is not just, as we would like to think, that we (...) choose more wisely, but also our luck in being the kind of people who respond in these ways. In this respect our attitude towards those who suffer (...) should not be ‘You asked for this’ but rather ‘There but for the grace of God go I’ ” (1998, p. 294).

On the face of it, there is nothing here for a proponent of the Potential Value of Option Set view to disagree with. For this view takes account of people’s “luck in being the kind of people who respond in certain ways to choices” by taking account of the ease or difficulty with which people can make certain choices. In the case of the Impetuous Woman, for example, the Potential Value of Option Set view assesses her option set under the Unclimbable Fence policy as more valuable for her than her option set under the Inform Everyone view, because it recognises the fact that in the latter case, she will be tempted to make a bad choice, while in the former case, she will not.

However, the expression “There but for the grace of God go I” suggests more than just the thought that our ‘luck’ in being the kind of people that we are plays *some* role in how we respond to our options. The source of this familiar expression is “There but for the grace of God goes John Bradford”, uttered by John Bradford (c.1510-1555), a Protestant preacher (later burned at the stake for his religious beliefs) on seeing a group of criminals being led to their execution. The remark sprung from Bradford’s commitment to a strict form of the doctrine of predestination, according to which man has no free will, and can be saved only by

the gift of God's grace, which he is powerless to earn through his choices.³² What the expression therefore suggests is that it is, in some sense, *wholly* a matter of luck that we respond to our options as we do.

Now, a defender of the Potential Value of Option Set view would, I think, find it difficult to accommodate this thought. For if our choices are, in this respect, on a par with our dispositions to choose, then why should one make allowances for the latter, but not the former in evaluating our option sets? However, belief in the truth of determinism seems to support the thought that how we respond to our choices is, in some sense, wholly matter of luck. There thus seems to be a conflict between the truth of determinism and the plausibility of the Potential Value of Option Set view.

The issue can also be put as follows. The Potential Value of Option Set view relies on the claim that a person can do otherwise than she does—hence, if determinism is accepted, differently than she will in fact be determined to do. So if we accept the Potential Value of Option Set view and determinism, we will have to develop some way of understanding the idea that a person “can do otherwise” that is compatible with determinism.

By contrast, the truth of determinism does not appear to undermine the form of justification the Realised Value of Option Set view offers for holding people substantively responsible.³³ For the Realised Value of Option Set view makes no reference to people's ability to choose otherwise than they are caused to do by deterministic causal laws. Instead, it appeals only to the way offering people a choice

³² As Bradford put it: “[It is the] Doctrine which demands our duty, but gives us no power thereto.”

See letter 67 ‘To certain men not rightly persuaded in the most true, comfortable, and necessary doctrine of God's holy election and predestination’ in *Writings of the Reverend John Bradford*, http://www.iclnet.org/pub/resources/text/ipb-e/epl-10/web/bradford_writings.html.

³³ My comments here parallel Scanlon's (1998, pp. 255-256) comments on the Value of Choice view.

realises the five values set out in section 3.1: the instrumental, expressive, and symbolic value of choice, the value of autonomy, and the value to others of giving a person a choice. None of these reasons appear to be sensitive to the truth of determinism; these values all seem equally worth securing whether or not our responses to our range of options all have ultimate deterministic causes outside us. What appears to be important for the first four of these values is that any such causes act ‘through us’, that is, they give rise to a set of judgements, aims, abilities, and character that exhibit the necessary coherence over time. So long as this is true, having a choice will sometimes be of instrumental value. Similarly, the mere fact that our capacities, judgements, and tastes have a causal source outside of us does not in itself make them any less our own; it therefore does not diminish the value of being given the opportunity to express them through our acts (expressive value), or of using them to actively shape, to some extent, our destiny through our own decisions (value of autonomy). Furthermore, being recognised by others as beings with the capacity, under certain circumstances, to make certain judgements and to act on them remains an important form of recognition (symbolic value). Finally, whatever causes others’ responses when they are faced with certain options, and whether or not a person’s judgements, aims, abilities, and character exhibit any coherence over time, the kind of response one may expect from them will sometimes make it valuable for one that they have certain options.

In sum, the truth of determinism would pose no challenge to the Realised Value of Option Set view, but pose significant challenges to the Potential Value of Option Set view. What if we possess libertarian freedom of the will—and determinism is therefore false? In that case, how we respond to our options is not always wholly a matter of luck, and the defender of the Potential Value of Option Set view can safely

appeal to the person's ability to do otherwise as a reason for giving bad outcomes that people could have avoided by choosing appropriately less weight in our moral thinking than similar outcomes that people could not have avoided. Indeed, suppose that we sometimes have an ability to act differently than we do, even if all the facts about the world prior to our actions—including our capacities, dispositions, and situation—are unchanged. Then it would seem not just justifiable, but also required, that we appeal to the value of options a person can achieve through his choice, rather than only to what he does achieve through his choice, in assessing his moral claims. In sum, I suggest that the relative attractiveness of the two views of substantive responsibility I have outlined may depend on the truth of determinism.

3.5 Brute luck equality and responsibility

I now wish to return to our discussion of Arneson's opportunity egalitarianism.

Arneson gives the following characterisation the view of responsibility that he sees as an essential part of his opportunity-egalitarianism:

“The idea of fair opportunity is that justice requires that a path be provided to each individual such that, if the individual stays on the path throughout her life, the outcome she reaches (...) would be fair. (...) But what happens to the individual if she strays off the path even by a slight amount is a ‘don't care’ from the standpoint of this conception of justice. (...) A young adult may behave in an irresponsibly careless way (...), then encounter incredibly bad luck and end up facing horribly grim life prospects that we could alleviate (...) at modest cost. The fair equality of opportunity [for] welfare account

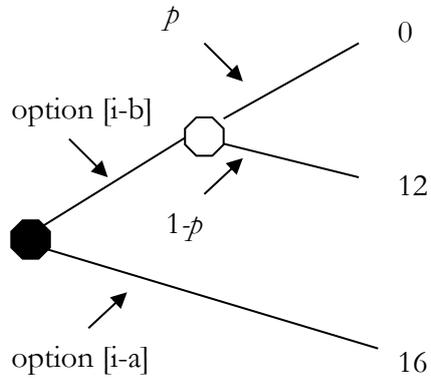
responds to such a case by insisting that justice demands no transfers of resources to alleviate the errant individual's plight because any such transfer would diminish the fair equality of opportunity to which others are entitled" (2001, p. 84).

The principle of responsibility which Arneson offers here appears to be the Forfeiture View's way of assessing the value of a person's option set, joined with the demand that we should choose among the class of equal option sets the option set which gives the highest value, thus determined. To recall, the Forfeiture View is that we should assess the value of opportunity sets by the value of just one accessible and relatively good path, which would be followed by a reasonably prudent person, and then choose which option set to give people by maximising the value of this path. This procedure involves ignoring the value of the other paths, and thereby the general impact on people's well-being of giving people this option set. On this view, we could give people an option set which would lead a substantial proportion to end up very badly, even though this could be prevented at relatively small cost by providing people with a different option set in which certain bad outcomes were far less bad.

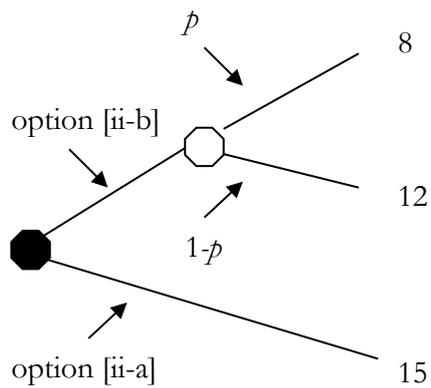
Arneson's example can be illustrated by figure 3.1. Imagine that we must choose between providing everyone with option set [i] or option set [ii]. As in chapter 2, black nodes denote a choice, white nodes a chance. The probability that the topmost path will be taken after a choice for option [a] is p . The numbers at the end of each path denote lifetime welfare. The difference between option sets [i] and [ii] then is that the value of the 'bad luck' path of option [b] is significantly increased at a small cost to the welfare of someone who takes the prudent path [a].

Figure 3.1 Two possible option sets

Option set [i]:



Option set [ii]:



Arneson's suggested principle of responsibility evaluates each of these option sets by the value of the prudent option only, and then requires us to choose the most valuable option set, which is option set [i].

Now, in section 3.2, we criticised the Forfeiture View of responsibility as paying insufficient attention to the value of eliminating disadvantageous options.

Interestingly, Arneson (2001, p. 84) agrees (see also Fleurbaey 1995a and Anderson

1999 for similar criticism).³⁴ Nonetheless, he sees his proposed principle of responsibility as an “ineliminable aspect of this ideal” of equal opportunity (2001, p. 84).

I believe that Arneson is mistaken that the Forfeiture View of responsibility is an ineliminable aspect of the ideal of equal opportunity. For one, interpreted as a merely an implication of the demands of brute luck equality, equal opportunity does not demand any inequalities due to choices at all. Of course, the impetus for the development of brute luck egalitarianism, as opposed to outright egalitarianism, has been to allow space for a view of responsibility that gives people’s options and choices a fundamental justificatory role. Therefore, Arneson’s ideal of opportunity egalitarianism is best understood as a combination of brute luck equality and such a view of responsibility. Now, we have seen that the Potential Value of Option Set view is a view of this kind, which avoids the unpalatable consequences of the Forfeiture view, because it assigns some weight to the value of each option in determining the value of an option set as a whole. Thus, the opportunity-egalitarian ideal could also be understood as requiring that we provide everyone with the most

³⁴ More than against the Forfeiture View, Fleurbaey’s and Anderson’s criticism is directed at a principle of responsibility which sees a normative role for the options and payoffs associated with these options that would arise in a laissez-faire economy with no market failures, and from general freedom to contract, against a background of brute luck equality. (As I mentioned in the Introduction of this thesis, a principle of this kind has been defended by Dworkin (2000), Eric Rakowski (1991) and Will Kymlicka (2002), and is also defended—though ultimately rejected—by Fleurbaey (1994, 1995a, 1995b, 1995c, 1998).) I have not discussed this principle here, in part because I believe that focusing on the value of people’s option sets, rather than assigning normative importance to essentially arbitrary no-intervention market payoffs, is a better way of developing a plausible principle of substantive responsibility (here I agree with Vallentyne 2002).

valuable equal option set, with this value being determined by the Potential Value of Option Set view.

In Arneson's example, suppose that we know that people are so constituted that, in spite of its being relatively disadvantageous, some significant share of people will choose option [b], either out of an impetuous conscious disregard for their own welfare, or simply because they pay insufficient attention to the likely consequences of choosing this option. This would merit giving option [b] some weight in assessing the value of the option set. Now, the value of option [b] will be larger in option set [ii] than in option set [i]. Therefore, if the weight given to option [b] in determining the overall value of an option set is sufficiently large, the Potential Value of Option Set view will judge option set [ii] to be better than option set [i].

In sum, I propose that we should see the opportunity-egalitarian ideal as characterised by a combination of brute luck equality and a principle which demands that we maximise the value of equal option sets as determined by the Potential Value of Option Set view. The discussion in section 3.4 suggests that the attractiveness of this ideal may depend on the truth of determinism. However, this version of the ideal is not vulnerable to the charge that it must permit choices through which people may end up very badly, and requires that we abandon people who make such choices, even when the costs of aiding them would be reasonable in comparison to the benefits such aid would bring them. It is not vulnerable to this criticism because it will count the elimination or improvement of disadvantageous options as valuable for people, and hence something that should be brought about when this is consistent with the demands of equality.

4. Equality, Responsibility, and Social Choice³⁵

Introduction

In the previous chapter, we concluded that the best combination of brute luck equality and responsibility requires that we provide everyone with the most valuable equal option set.³⁶ In this chapter, I will examine, first, how we can represent this requirement in social choice rules, and second, which social choice rule we should use if we replace brute luck equality by absolute priority to those who are most disadvantaged by brute luck. I will do so with the help of a hypothetical example of returns to education for individuals from different ethnic groups.

³⁵ This chapter draws on material from my joint paper with Matthias Hild, 'Equal Opportunity and Opportunity Dominance', which appeared in *Economics and Philosophy* 20 (2004), pp. 117-145. The material from that paper has been substantially revised here. I thank Matthias Hild for the fruitful collaboration on that paper, and the following people for comments on that paper as it evolved: Brian Barry, Ken Binmore, Alex Brown, Jerry Cohen, Marc Fleurbaey, Dirk Van de gaer, Jeroen Knijff, Peter Postl, Michael Otsuka, John Roemer, Robert van der Veen, Peter Vallentyne, Jonathan Wolff, and an anonymous referee for *Economics and Philosophy*. Earlier versions of this chapter were presented at the Analytical Philosophy National Postgraduate Conference in Reading (May 2000), the Workshop on Equal Opportunity at the University of Bayreuth (February 2001), The Political Theory Seminar at Yale University (May 2002), the Meeting of the Society for Social Choice and Welfare at Caltech (June 2002), and the Economics Seminar at London Metropolitan University (May 2003). I thank the participants in these meetings for their comments.

³⁶ Taking, as mentioned in chapter 3, the Potential Value of Option Set approach to evaluating an individual's option set.

Before introducing this example, I cover some preliminaries. In section 4.1, I introduce some terminology from the relevant social choice literature. In section 4.2, I review solutions in the social choice literature to two difficult problems that typically need to be solved before responsibility-sensitive egalitarian social choice rules can be applied to real-world cases. The first problem is how we can infer from observable data the range, accessibility and quality of the options open to individuals, and these individuals' ability to negotiate these options. I review John Roemer's well-known solution to this problem.³⁷ I argue that Roemer's solution is flawed because it wrongly equates the range of the options that are open to an individual with the range of options chosen by other individuals who are placed in similar circumstances.

The second problem is how to determine the value of an option set. I review Dirk Van de gaer's solution to this problem, which is to equate the value of an individual's option set with the average value achieved by other individuals who are placed in similar circumstances.³⁸ I argue that Van de gaer's method is flawed because it ignores the possibility that the value of an option set that members of a group face can differ from the average value that members of that group achieve through their choices.

In section 4.3, I introduce the central example of the chapter, and discuss the implications of providing everyone with maximally valuable equal option sets.

³⁷ See Roemer (1993, 1996, 1998, 1999, 2002, 2003). For applications of Roemer's proposal to practical cases, see Llavador and Roemer (2001), Betts and Roemer (1999), Roemer *et al* (2003), and Van de gaer, Schokkaert, and Martinez (2001).

³⁸ See Van de gaer (1993), Bossert, Fleurbaey and Van de gaer (1999), and Van de gaer, Schokkaert, and Martinez (2001), and Ooghe, Schokkaert, and Van de gaer (forthcoming).

I then turn to the question of which social choice rule we should use if we replace brute luck equality by absolute priority to those who are most disadvantaged by brute luck. Two ways of determining who is most disadvantaged by brute luck have been discussed in the responsibility-sensitive egalitarian social choice literature. One way is to take the most disadvantaged to be all individuals who are least well off compared to others who have chosen comparable options (see Roemer (1993, 1996, 1998, 2002, 2003)).³⁹ For example, if we are considering returns to education, then we might count the individuals who have the lowest returns to a given educational choice (say, dropping out of high school) as among those who are most disadvantaged by brute luck. For others who chose a similar option (e.g. dropping out of high school) are better off due to no choice or fault of their own, and this is a matter of brute luck. I will call the approach that aims to improve the situation of the least well off, so defined, the Leximin Value of Options approach.

A second way is to start by assessing the value of individuals' option sets as a whole, and then take those who are most disadvantaged by brute luck to be all individuals with the least valuable option set (see Van de gaer 1993). I will call the approach that aims to improve the situation of the least well off, so defined, the Leximin Value of Option Set approach.

In a recent paper, Erwin Ooghe, Erik Schokkaert, and Dirk Van de gaer (forthcoming) carefully analyse particular versions of these two approaches and the differences between them, but do not argue in favour of one of these approaches.

³⁹ For completeness, we should add: when this being worse off than others is not due to the outcome of the option being uncertain, since in that case an individual might be worse off than another who made the same choice without being disadvantaged due to brute luck. For simplicity, I will assume the absence of uncertainty in what follows.

This chapter supplements their work, by further examining both approaches, and by arguing in favour of the Leximin Value of Option Set approach.

In sections 4.4-4.6, I develop the Leximin Value of Options approach. In section 4.4 I develop a social choice rule, which I call ‘option-dominance’, which gives a partial ordering of policies by de-selecting all policies that are uncontroversially worse than some other policy on this approach.

In section 4.5, I discuss the relationship between Roemer’s social choice rule and the Leximin Value of Options View. I argue that an adjusted form of Roemer’s rule can be used to select a single policy from the set of option-undominated policies.

In section 4.6, I discuss the relationship between option-dominance and the Principle of Personal Good (or PPG for short). The PPG—that for all alternatives ϕ and ψ , if everyone is at least as well off under ϕ as under ψ , and someone is strictly better off, then ϕ is better than ψ ⁴⁰—is the core of welfare economics. The PPG, however, is not well suited to situations in which individuals’ responsibility for their welfare plays a role in assessing their claims. For the PPG is concerned with individuals’ welfare outcomes, whereas once we take account of individuals’ responsibility for their welfare, our attention shifts from their welfare outcomes to the quality of their options. It is therefore interesting to examine the relationship between responsibility-sensitive social choice rules and the PPG, to see whether the latter principle must be abandoned when we take account of individuals’ responsibility for their welfare.

⁴⁰ This principle, which was introduced by Broome (1991), is structurally equivalent to the strong Pareto principle, except that it is not formulated in terms of preferences. Since, in chapter 1, I have argued in favour of a non-preference-based conception of welfare, the PPG is more appropriate for my discussion than the Pareto-principle.

I show that while option-dominance and the PPG are independent of each other, in the sense that one of them can express a strict preference where the other remains silent, they cannot contradict each other, i.e. it cannot be the case that a first policy dominates a second by the PPG while the second policy option-dominates the first.

In section 4.7, I show that though it aids those disadvantaged by one form of brute luck—being worse off than others who make similar choices—the Leximin Value of Options approach neglects another form of brute luck: being disadvantaged because one’s option set is less valuable than others’ option sets, either because one does not have good options that others do have, or because, while one has the same options that others have, the outcomes of these options are all relatively low.

In section 4.8, I argue that under the assumption that individuals can choose among their options under adequate conditions of choice, brute luck inequality in the value of individuals’ option sets is more important than brute luck inequality in the value of individuals’ options. I conclude that the Leximin Value of Option Set approach is a better way of integrating priority for those who are disadvantaged by brute luck with the demands of responsibility.

The main argument of the chapter can be followed without recourse to the more formal arguments, which I have placed in sub-sections marked by a *. Calculations and proofs that are not of central importance appear in the endnotes.

4.1 Equal option sets and social choice

Equal option sets and social choice 1

Let us first introduce some terminology from the relevant social choice literature. In this literature, it is common practice to define the issues in the following way. We start by listing all the factors that influence individuals' achievement of welfare. We then sort these factors into three categories. The first category consists of those factors that are under control of the institution or institutions that regulate the distribution of welfare. Call these 'policy variables'. The second category consists of those factors that are the result of choices for which individuals can be held responsible. For each individual, these factors will take on certain values, which I will call individuals' 'responsible characteristics'. The third category consists of those factors for which individuals cannot be held responsible. For each individual, these factors will take on certain values, which I will call individuals' 'non-responsible characteristics'. Individuals then have equal option sets just in case the policy variables are chosen so that (a) individuals with the same responsible characteristics attain the same welfare outcomes, irrespective of their non-responsible characteristics; (b) the accessibility of each responsible characteristic is independent of individuals' non-responsible characteristics; (c) individuals have an equal and adequate ability to negotiate the choices that lead to them acquiring these responsible characteristics; and (d) individuals have equal and adequate information.

** Equal option sets and social choice 2*

Somewhat more formally, we can express this definition as follows (see also Bossert 1995 and Fleurbaey 1995a, p. 30). Let $Y = (Y_1, \dots, Y_M)$ be a vector of factors that can influence individuals' level of welfare for which individuals are responsible and $Z = (Z_1, \dots, Z_N)$ be a vector of factors that can influence individuals' level of welfare for which individual cannot be held responsible. Let Φ be a set of feasible policies. Then $w(\phi, y, z)$ measures the attainment of welfare under policy $\phi \in \Phi$ by an individual with a combination $Y = y$ of characteristics for which she can be held responsible and a combination $Z = z$ of characteristics for which she cannot be held responsible.

A policy ϕ ensures that individuals have equal option sets just in case (a) $w(\phi, y, z) = w(\phi, y, z')$ for any combination of responsible characteristics y and any combinations of non-responsible characteristics z and z' (as long as some individual displays (y, z) and (y, z')); (b) the accessibility of each responsible characteristic is independent of individuals' non-responsible characteristics (c) individuals have an equal and adequate ability to negotiate the choices that lead to them acquiring these responsible characteristics; and (d) individuals have equal and adequate information.

4.2 Identifying and evaluating individuals' option sets

Roemer on 'relative effort'

In real-world cases, we may have to rely only on observable characteristics and choices of individuals, from which we have to infer the range and quality of the options open to them, their ability to negotiate these options, and the relative accessibility of these options. This is obviously a complex matter; nonetheless, it typically must be solved before responsibility-sensitive egalitarian social choice rules can be applied to concrete cases. Roemer (1993, 1996, 1998) proposes the following way of identifying individuals' option sets and of determining that these are equal. He selects some measure of individuals' non-responsible characteristics, and sorts individuals into 'types' on the basis of these characteristics, with individuals with the same or very similar non-responsible characteristics falling into the same type. He then selects some choice variable that is an important contributor to the outcome for which we want to establish equal opportunity. Roemer calls this variable individuals' 'effort'.

Now, by construction, everyone in the same type will have equal option sets. The question is how we can ensure that individuals from different types have equal option sets. Roemer points out that ensuring that individuals in different types who make the same effort get the same welfare is not, as a rule, sufficient for ensuring that they have equal option sets. For individuals in different types may have a different ability to negotiate their options, or find it more or less difficult to achieve a given level of effort. Roemer proposes to compensate for these differences as follows.

He first ranks all individuals within a type according to their effort. This ranking allows him to partition individuals within each type into quantiles of effort relative to their type. (A person i belongs to quantile π of relative effort exactly if a fraction π of the entire population expends at most as much effort as i .) Roemer calls an individual's effort quantile this individual's 'relative effort'.⁴¹ Roemer assumes that for any level of relative effort π , the ease or difficulty with which an individual can choose that level of relative effort is independent of his type. Moreover, he assumes all individuals are equally able to negotiate the choice of a level of relative effort. It follows that individuals have equal option sets when all individuals who expend the same degree of relative effort have equal outcomes (Roemer 1998, pp. 15-16).

We can illustrate this procedure and one of its flaws in the following example, which I will call the Smoking Case.⁴² Suppose that at time $t=0$ a group of individuals are all of the same type, and all have equal option sets. For all individuals, welfare is an identical function of government tax policy and one choice variable, the number of cigarettes they smoke per day, with welfare decreasing in the number of cigarettes smoked. The population is uniformly distributed over the range of smoking 0-20 cigarettes per day. At time $t=1$, with the aim of increasing sales, a cigarette manufacturer directs advertising at the half of the population that smokes between 0-10 cigarettes a day. Suppose that only and all these individuals are exposed to the advertising campaign, and that these individuals cannot avoid exposure to the

⁴¹ In later work, Roemer switches to a different measure of relative effort, which he equates with the quantile that an individual occupies in his type's distribution of welfare outcomes (see Roemer 2002, 2003). The criticism expressed in the main text applies equally to this method of constructing relative effort.

⁴² For further criticism of Roemer's method see Fleurbaey (1998).

campaign. This campaign makes cigarettes more attractive to the exposed individuals; as a result, they start smoking more than before, and are uniformly distributed over the range of 5-15 cigarettes per day.

In the Smoking Case, one should conclude that individuals who were exposed to the advertising campaign face a worse option set than before, as it has become more difficult for them to refrain from smoking, and that they therefore have a worse option set than non-exposed individuals. As we will see, however, Roemer's method draws the opposite conclusion. On his approach, being exposed is a non-responsible characteristic, and hence grounds for sorting the individuals who formerly belonged to one type into two types. Let us call them the 'exposed' and 'non-exposed' type. Roemer's method would have us equalise the welfare of individuals in the non-exposed and exposed types who are at the same relative position in their type's distribution of smoking behaviour. This would mean equalising the welfare of an exposed individual who smokes 5 cigarettes a day with the welfare of a non-exposed individual who smokes 10 cigarettes a day, and so on. This would involve taxing exposed individuals and transferring the receipts to non-exposed individuals. Roemer's method thus implies that exposure gives individuals an unfair advantage, which needs to be eliminated by government policy in order to ensure equality of individuals' option sets.

Roemer's method goes wrong in this case because it equates the range of options open to individuals within a given type with the range of choices that members of that type actually make. In this case, it therefore fails to identify that, by hypothesis, the unexposed individuals had the option of not smoking at all, even though they all chose to smoke.

Van de gaer (1993) proposes a related method of assessing the quality of individuals' option sets. Van de gaer also sorts individuals into types on the basis of their non-responsible characteristics. He then takes all individuals with the same type to have the same option set, and equates the value of the option set with the average welfare achieved by individuals who have this option set.⁴³

Van de gaer's method is open to a similar objection as Roemer's method. In the Smoking Case, Van de gaer's method mistakenly takes the value of the option set faced by the exposed individuals to be greater than the value of the option set faced by the non-exposed individuals, simply because the latter choose to smoke more, and hence have a lower average level of welfare.

Of course, Roemer's and Van de gaer's methods are not altogether misguided; in many real-world cases, the actual pattern of choices and outcomes exhibited by individuals of a certain type may be a good first guide to identifying the options open to them and the relative accessibility of these options. However, it is not more than a first guide, and so long as we believe that two groups of individuals all of whom have

⁴³ Ensuring that individuals have an equally valuable option set in Van de gaer's sense is not a sufficient condition for ensuring that they have equal option sets. For two individuals with an equal ability to negotiate their options, and equally accessible options, might have equally valuable option sets, in Van de gaer's terms, because the value of the first individual's best option is much greater than the value of the second individual's best option, and the value of the first individual's second-best option is much less than the value of the second individual's second-best option. This means that what they can achieve is still, to some extent, a matter of brute luck. Nonetheless, assessing the value of individuals' option sets might be a good alternative to making them equal when the latter is not feasible; and we will see that it will be a useful approach when we opt for leximin over strict equality.

equal option sets might exhibit different overall patterns of choice, we cannot regard all differences in the pattern of choices and outcomes between these groups as reflecting differences in the quality and/or accessibility of their options. We cannot, therefore, infer the quality and accessibility of the options faced by a group of individuals from the pattern of choices made and outcomes achieved by that group in the simple ways proposed by Roemer and Van de gaer.

4.3 Equal option sets in an example

An example

In the next few sections, I will illustrate various social choice rules with the help of a hypothetical example of the relation between schooling, ethnicity, and income. The ‘status quo situation’ is represented in table 4.1.⁴⁴ As the table makes clear, in our hypothetical country, Nationals earn a substantially higher mean income than members of immigrant minorities with the same educational qualifications. Those born into Immigrant Group 1 and Group 2 alternate in having the lowest mean income at a given educational level. Those born into Group 2 have the highest mean

⁴⁴ The numbers in this table are in fact drawn from the U.S. Census Bureau’s database, and reflect the relation between mean income, schooling, and race in the U.S.A. in 1999. However, the many simplifying assumptions I make for the purpose of our discussion mean the example and our analysis of it are entirely unconnected to any actual policy or political issues, and nothing in what follows should be taken to suggest the desirability of any policy. To avoid any suggestion of an actual relation of this kind, I have used invented labels for the original racial categories, which were ‘White’, ‘Black’, and ‘Other’.

level of schooling, while those born into Group 1 have the lowest mean level of schooling.

Table 4.1 Median income by ethnicity and education

	High School No Degree	High School w/Degree	College No Degree	Associate Degree	BA or more	Total
NATIONALS						
number ('000)	23,816	52,642	31,574	12,218	39,338	159,588
percentage	12.5%	27.7%	16.6%	6.4%	20.7%	84.0%
income	14,885	23,822	27,930	32,116	54,208	31,426
IMMIGRANT GROUP 1						
number ('000)	4,655	7,581	4,812	1,401	3,398	21,847
percentage	2.5%	4.0%	2.5%	0.7%	1.8%	11.5%
income	11,948	19,934	24,445	29,115	42,361	23,306
IMMIGRANT GROUP 2						
number ('000)	1,299	1,985	1,541	598	3,093	8,516
percentage	0.7%	1.0%	0.8%	0.3%	1.6%	4.5%
income	13,129	20,199	22,330	28,769	49,966	30,919
TOTAL						
number ('000)	29,770	62,208	37,927	14,217	45,829	189,951
percentage	15.7%	32.7%	20.0%	7.5%	24.1%	100.0%
income	14,349	23,232	27,260	31,684	53,043	30,469

Now, in any real-world case, the causes of a pattern of educational achievement and mean income of this kind would obviously be complex, and include factors such as parental income and education, the quality of the educational institutions attended, individuals' abilities, and their preferences and information.

For the purposes of our discussion, however, I will make the following simplifying assumptions. First, that an individual's level of welfare is equivalent to his income. Second, that all individuals at a particular level of education achieve the mean income in of their ethnic group at that level of education. Third, that all individuals have the same set of options, consisting of the five educational levels, and that—differences in the distribution of individuals from different ethnic groups across different educational levels notwithstanding—all these options are equally accessible to all individuals, independently of their ethnic group. Fourth, that all

individuals have an adequate and equal ability to negotiate their options, and adequate and equal information. These assumptions together imply that all individuals have equal option sets when the income associated with any given level of education is independent of individuals' ethnicity.

(In the terminology introduced in section 4.1, this example can be expressed as follows: individuals' welfare is a function of government policy, a 'responsible factor', education, and a 'non-responsible factor', ethnicity. Individuals' sole 'responsible characteristic' is their level of education, and their sole 'non-responsible characteristic' is their ethnicity.)

With respect to government policy, I assume government can intervene only by redistributing income, which has a constant variable cost. Thus, for each unit of income collected, only ρ units are available for redistribution while the remainder $1 - \rho$ covers the cost of the intervention. For concreteness, I will assume that this cost amounts to 0.2 per currency unit redistributed, so that $\rho = 0.8$.

Finally, unless otherwise specified, I will assume that transfers do not affect individuals' choices of an educational level (i.e. there are no disincentive effects of taxation).

Determining the value of option sets

Unfortunately, to my knowledge, the social choice literature contains no satisfactory solution to the problem of evaluating a person's option set.⁴⁵ Nor do I have a solution of my own. For illustrative purposes only, I will therefore assume the

⁴⁵ For some introductory discussion and references see Sen (1999, pp. 75-79 and 2002, especially chapters 20-22).

following very simple method of determining the value of an option set: the value of option set of someone from a particular ethnic category is a weighted average of the value of the five options she faces—that is, of the income associated with each of the five educational levels—with the weights given by the share of the total population who choose each educational level.⁴⁶ Since the total population is distributed differently across different educational choices in different ethnic categories, this means the value of the option set of an individual from a particular ethnic category will differ from the average value that individuals from that ethnic category achieve through their choices, as illustrated by table 4.2. The difference between the two reflects the degree to which individuals in a given ethnic category choose, on average, higher or lower levels of education than the mean of the total population.

Table 4.2 Value of option sets versus mean income by ethnicity in status quo

Ethnicity	Value of option set in status quo	Mean income in status quo	Difference
NATIONALS	31,186	31,426	240
IMMIGRANT GROUP 1	25,676	23,306	-2,370
IMMIGRANT GROUP 2	27,332	30,919	3,587

⁴⁶ Since I have assumed that all these options are equally accessible to all individuals, independently of their ethnicity, and that all individuals have an equal ability to negotiate their options, it is appropriate that the weights do not depend on the pattern of choice within each ethnic category, but are identical across the population. Equating the weight given to a particular educational level with the share of the overall population choosing that level was done because it can be seen as an—admittedly very rough—first indicator of the relative accessibility of these options. My conclusions about the Leximin Value of Option Sets approach do not depend on this simple way of judging the quality of an option set.

Equal option sets in our example

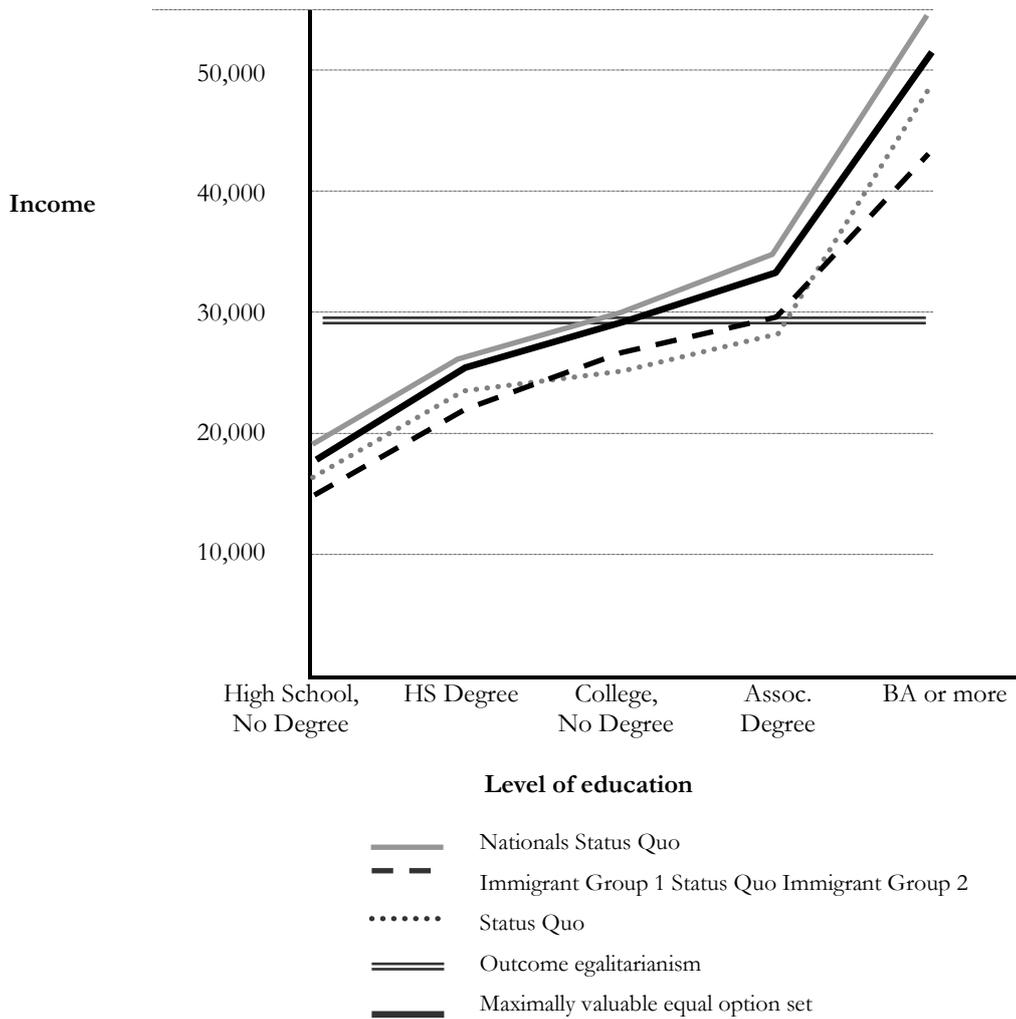
Given our assumptions, if we consider income redistributions on the basis of both ethnicity and education, it is possible to achieve equal option sets for all individuals. Indeed, we can do so in many different ways. Table 4.3 shows two such ways.¹ One way is to equalise the income of all individuals at the highest possible level. This is outcome egalitarianism, which is, of course, consistent with the demands of brute luck egalitarianism. However, once we bring in the demands of substantive responsibility as well, we choose the policy which gives people the most valuable equal option set. This yields the second policy listed in table 4.3. These policies are also compared with the status quo distribution and with each other in figure 4.2.

Two things are notable in this figure and table. The first is the role played by our theory of substantive responsibility in justifying inequalities. Whereas brute luck equality alone is consistent with many different patterns of reward to education, including equality of outcome, the demand to maximise the value of equal option sets, in the context of costly redistribution, leads us to a distribution which, though it contains no unchosen inequalities, does yield significant inequalities due to different choices of education levels. The second is the fact that giving equal option sets need not eliminate inequalities in the distribution of income between ethnic groups: so long as everyone's option sets are equal, inequalities between groups that are due to different patterns of choice between these groups are entirely legitimate.

Table 4.3 Two ways of giving people equal option sets

	High School No Degree	High School w/Degree	College No Degree	Associate Degree	BA or more	Total
Outcome egalitarianism						
NATIONALS income	29,888	29,888	29,888	29,888	29,888	29,888
change	14,403	5,466	1,358	-2,828	-24,920	-2,138
IMMIGRANT GROUP 1 income	29,888	29,888	29,888	29,888	29,888	29,888
change	17,340	9,354	6,958	133	-13,073	5,982
IMMIGRANT GROUP 2 income	29,888	29,888	29,888	29,888	29,888	29,888
change	16,159	9,089	6,958	519	-20678	-1,632
TOTAL income	29,888	29,888	29,888	29,888	29,888	29,888
change	14,439	6,055	2,027	-2,396	-23,756	-1,182
Maximise value of equal option set						
NATIONALS income	14,247	23,113	27,127	31,594	52,802	30,551
change	-638	-709	-803	-522	-1,406	-875
IMMIGRANT GROUP 1 income	14,247	23,113	27,127	31,594	52,802	27,269
change	2,299	3,179	2,682	2439	10,441	3,964
IMMIGRANT GROUP 2 income	14,247	23,113	27,127	31,594	52,802	33,865
change	1,118	2,914	4,797	2825	2,836	2,946
TOTAL income	14,247	23,113	27,127	31,594	52,802	30,551
change	-102	-120	-134	-90	-241	-147

Figure 4.2 Returns to education in the status quo, and under two brute-luck egalitarian policies.



4.4 Option dominance

I now turn to the question which social choice rule we should use if we replace brute luck equality with absolute priority to those who are most disadvantaged by brute luck. As mentioned, one way of determining who is most disadvantaged by brute luck is to compare different individuals who have chosen comparable options. In our example, this means comparing all individuals at a given education level, and

determining which individuals are least well off at that education level. We can see in table 4.1 and figure 4.2 that in the status quo, those born into Immigrant Group 1 are worst off at education levels ‘High School, No Degree’, ‘High School Degree’, and ‘BA or more’. Those born into Immigrant Group 2 are worst off at education levels ‘College, No Degree’, and ‘Associate Degree’. The reason for focusing on each option separately is that this captures the following complaint on the part of someone from Immigrant Group 1 who drops out of high school: ‘My ethnicity it is a matter of brute luck; if I had been a National or a member of Immigrant Group 2 and made the same choice, I would have been better off. I am therefore among those who are most disadvantaged by brute luck’. As mentioned, I will call the approach that aims to improve the situation of those who are disadvantaged by brute luck in this way the Leximin Value of Options approach.

On this approach, then, we should give priority to improving the situation of all those individuals who are less well off compared to others who have made similar choices. In terms of evaluating policies, this leads to the following criterion of betterness for each education level:

- A policy ϕ is *better than* a policy ψ *conditional on* a particular education level y exactly when the worst off individuals at education level y under ϕ are better off than the worst off individuals at education level y under ψ and, in the case of a tie, the second worst off individuals at education level y under ϕ are better off than the second worst off individuals at education level y under ψ , etc.
- A policy ϕ is *as good as* ψ *conditional on* y exactly when the worst off individuals at y under ϕ are as well off as the worst off individuals at y

under ψ and the second worst off individuals at y under ϕ are as well off as the second worst off individuals at y under ψ , etc.

- A policy ϕ is *at least as good as ψ conditional on y* exactly if ϕ is as good as or better than ψ conditional on y .

This criterion only tells us whether one policy is better than another for a given level of education. Policies may not be best across all levels of education: a policy that is better than the status quo for the least well off high school dropouts might be worse than the status quo for the least well off college graduates. For example, if we confine our attention to a comparison between the outcome-egalitarian policy and the status quo in figure 4.2, we can see that neither is uniformly better than the other across all levels of education.

However, in some comparisons, one policy may be better than another across all education levels. For example, if we compare only the ‘maximise value of equal option set’ policy of table 4.3 with the status quo, then we can see that the former is better than the status quo at every education level. The following criterion of *option-dominance* captures this idea of uniform superiority:

- A policy ϕ *dominates* a policy ψ in the *value of options*, or *option-dominates* ψ , if and only if ϕ is at least as good as ψ conditional on every education level attained by some individuals and better conditional on some education level.

Suppose we now use this criterion to eliminate all policies that are option-dominated. We then typically get an incomplete ranking of policies. All policies that are uncontroversially worse than some other policy from the perspective of the Leximin Value of Options approach will be eliminated, but we may be left with many policies that do not option-dominate each other. For example, if we confine

ourselves to the three policies represented in figure 4.2, we can see that the criterion of option-dominance eliminates the status quo, because it is dominated by the ‘maximum value of option set’ policy, but does not eliminate the outcome-egalitarian policy. That is because the outcome-egalitarian policy is better than the maximum value of option set policy for some education levels, but worse for others.

* *A formal definition of option-dominance*

As before, let $Y = (Y_1, \dots, Y_M)$ be a vector of responsible factors that can influence individuals’ welfare and $Z = (Z_1, \dots, Z_N)$ be a vector of non-responsible factors that can influence individuals’ welfare. Let Φ be a set of feasible policies. Then $w(\phi, y, z)$ measures the welfare under policy $\phi \in \Phi$ by an individual with a combination $Y = y$ of responsible characteristics a combination $Z = z$ of non-responsible characteristics.

Let $\phi, \psi \in \Phi$, let N_y^ϕ be the number of y -individuals under policy ϕ and let N_y^ψ be the number of y -individuals under policy ψ , allowing for behavioural responses to policies that change responsible characteristics. Number the y -individuals under ϕ by $1, \dots, N_y^\phi$, and number the y -individuals under ψ by $1, \dots, N_y^\psi$.

Let σ be a permutation of $\{1, \dots, N_y^\phi\}$ such that

$w(\phi, y_{\sigma(1)}, z_{\sigma(1)}) \leq \dots \leq w(\phi, y_{\sigma(N_y^\phi)}, z_{\sigma(N_y^\phi)})$ and let τ be a permutation of

$\{1, \dots, N_y^\psi\}$ such that $w(\psi, y_{\tau(1)}, z_{\tau(1)}) \leq \dots \leq w(\psi, y_{\tau(N_y^\psi)}, z_{\tau(N_y^\psi)})$.

- Then ϕ is *better than* ψ conditional on $Y=y$ exactly when there exists some

$1 \leq i \leq \min(N_y^\phi, N_y^\psi)$ such that $w(\phi, y_{\sigma(i)}, z_{\sigma(i)}) > w(\psi, y_{\tau(i)}, z_{\tau(i)})$ and,

for all $1 \leq j \leq i$, $w(\phi, \mathcal{Y}_{\sigma(j)}, \mathcal{Z}_{\sigma(j)}) = w(\psi, \mathcal{Y}_{\tau(j)}, \mathcal{Z}_{\tau(j)})$. This strict relation is asymmetric and transitive.

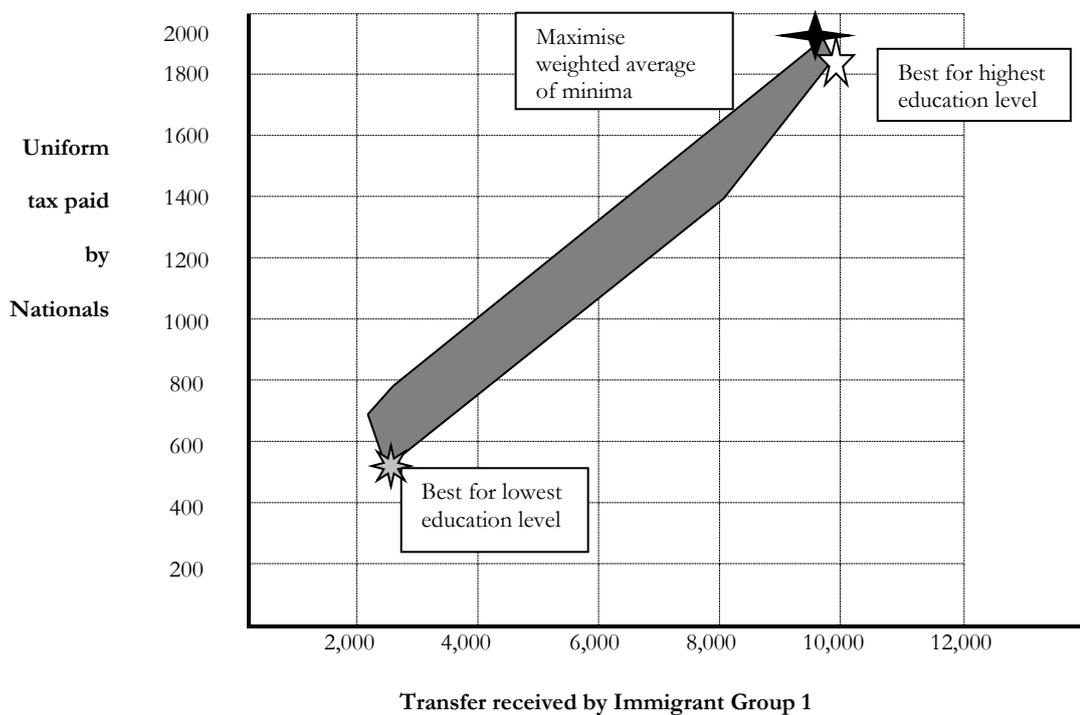
- When $N_y^\phi = N_y^\psi$ and ϕ and ψ are not better than each other conditional on y , we say that ϕ is *as good as* ψ conditional on $Y=y$, but refrain from such comparisons for policies that differentially affect the number of individuals with characteristic y .
- We say that ϕ is *at least as good as* ψ conditional on $Y=y$ exactly if ϕ is as good as or better than ψ conditional on $Y=y$. This relation is transitive, but not complete.
- We say that ϕ *option-dominates* ψ exactly if ϕ is at least as good as ψ conditional on every value $Y=y$ (displayed by some individual) and ϕ is better than ψ conditional on some value $Y=y'$ (displayed by some individual). Option-dominance is transitive, but not complete.
- Finally, ϕ is *option-undominated*, or *option-optimal*, just in case there exists no $\psi \in \Phi$ that option-dominates ϕ . If there are only finitely many feasible policies, then the set of option-undominated policies is non-empty.

Option-dominance further illustrated

In the policy environment that we assumed at first, in which the government can redistribute on the basis of both ethnicity and education, there are no changes in individuals' choice of education level, and transfers have a constant cost, applying the leximin rule for each level of education just leads to equality at each level of education. To illustrate the option-dominance rule in cases where it does not select

only policies that provide everyone with equal option sets, let us now consider redistribution on the basis of ethnicity alone. This means that every member of an ethnic group pays the same amount in tax, or receives the same transfer. An application of our principle of option-dominance then selects the shaded area in figure 4.3.ⁱⁱ

Figure 4.3 Option-undominated policies with transfers based on ethnicity



This again illustrates that option-dominance typically selects a set of policies rather than a single policy. Of any two option-undominated policies ϕ and ψ , it will be true that either (i) ϕ is better than ψ conditional on some level of education, and worse than ψ on some other level of education, or (ii) ϕ is just as good as ψ across all levels of education, or (iii) ϕ and ψ are incomparable on some level of

education.⁴⁷ Case (i) is illustrated by the two option-undominated policies marked in figure 4.3 and represented in table 4.4: the ‘best for lowest education level’ policy is the one that makes the worst off high school dropouts as well off as possible, given the set of feasible policies. By contrast, the ‘best for highest education level’ policy makes the worst off college graduates as well off as possible.

Since the Leximin Value of Options approach typically does not select a single policy, we need a further criterion to choose between option-undominated policies. We will examine one possible criterion of this kind in the next section.

⁴⁷ It follows from our definitions that two policies ϕ and ψ will be incomparable on a given level of education y if either (i) under ϕ , there are no individuals who have chosen education level y , whilst under ψ some individuals do choose education level y ; or (ii) if there are N_y^ϕ y -individuals under ϕ and N_y^ψ y -individuals under ψ , and the least well off, the second least well off, ... the N_y^ϕ -th least well off y -individuals under ϕ and ψ are equally well off, and $N_y^\phi < N_y^\psi$.

Table 4.4 Two option-undominated policies

	High School No Degree	High School w/Degree	College No Degree	Associate Degree	BA or more	Total
	Best for lowest education level					
NATIONALS income change	14,384 -501	23,321 -501	27,429 -501	31,615 -501	53,707 -501	30,925 -501
IMMIGRANT GROUP 1 income change	14,384 2436	22,370 2436	26,881 2436	31,591 2436	44,797 2436	25,742 2436
IMMIGRANT GROUP 2 income change	14,384 1,255	21,454 1,255	23,585 1,255	30,024 1,255	51,221 1,255	32,174 1,255
TOTAL income change	14,383 35	23,146 -87	27,203 -57	31,546 -138	52,879 -165	30,385 -84
	Best for highest education level					
NATIONALS income change	13,018 -1867	21,955 -1867	26,063 -1867	30,249 -1867	52,341 -1867	29,559 -1867
IMMIGRANT GROUP 1 income change	21,928 9981	29,914 9981	34,425 9981	39,096 9981	52,341 9981	33,286 9981
IMMIGRANT GROUP 2 income change	15,503 2,374	22,573 2,374	24,704 2,374	31,143 2,374	52,341 2,374	33,293 2,374
TOTAL income change	14,548 199	22,948 -284	26,924 -336	30,707 -997	52,341 -702	30,156 -313

4.5 Option-dominance and Roemer's social choice rule

When stripped of the 'relative effort' metric we criticised in section 4.1, Roemer's (1993, 1996, 1998, 2002, 2003) social choice rule suggests a way of selecting from the set of option-undominated policies. For Roemer's approach is very close to the Leximin Value of Options approach. Conditional on each combination of responsible characteristics y , Roemer uses the maximin relation of betterness instead of the leximin relation. Unlike leximin, the maximin relation assigns a numerical value to a policy for each responsible characteristic, viz., the welfare of the least well off individuals at y . He then maximises a weighted average of these values, with the welfare of the least well off individuals at the combination of responsible characteristics y being given the weight of the number of individuals that are least well off at that combination of responsible characteristics. If N is the number of individuals in the population, $N_{y\bar{z}}$ is the number of individuals with the combination of responsible characteristics y and non-responsible characteristics \bar{z} , and $w(\phi, y, \bar{z})$ measures the welfare of an individual with responsible characteristics y and non-responsible characteristics \bar{z} , then Roemer selects a policy ϕ that maximises:

$$\sum_y \min_{\bar{z}} \frac{N_{y\bar{z}}}{N} \cdot w(\phi, y, \bar{z})$$

In the context of our example, this means maximising the average welfare of all individuals who are least well off at each level of education.

Roemer's rule has the drawback that it might select a policy that is option-dominated, for the following reasons. First, because it ignores the welfare of individuals who are better off than the least well off at each responsible characteristic. Second, because the weighting scheme favours policies that reduce the number of individuals that are least well off at a relatively unrewarding combination of responsible characteristics (e.g. high school dropouts) in comparison with the number of individuals that are least well off at a relatively rewarding combination of responsible characteristics (e.g. college graduates).ⁱⁱⁱ However, this shortcoming can be remedied by using it to select from the set of option-undominated policies. Used in this way, Roemer's rule provides one possible way of selecting from this set, viz. by maximising a weighted average of the welfare of the least well off for each combination of relevant characteristics. When transfers are based on ethnicity alone, the resulting policy is the 'maximise weighted average of minima' policy pictured in figure 4.3, and characterised in table 4.5.^{iv}

Table 4.5 Roemer's policy

	High School No Degree	High School w/Degree	College No Degree	Associate Degree	BA or more	Total
	Maximise weighted average of minima					
NATIONALS income change	12,945 -1940	21,882 -1940	25,990 -1940	30,176 -1940	52,268 -1940	29,486 -1940
IMMIGRANT GROUP 1 income change	21,855 9,907	29,841 9,907	34,352 9,907	39,062 9,907	52,268 9,907	33,213 9,907
IMMIGRANT GROUP 2 income change	16,789 3,660	23,859 3,660	25,990 3,660	32,429 3,660	53,626 3,660	34,579 3,660
TOTAL income change	14,506 157	22,915 -318	27,051 -209	31,416 -537	52,360 -648	30,143 -326

This weighted average of the welfare of the least advantaged provides one natural way (but not the only way) of selecting from the set of option-undominated policies.

4.6 Option dominance and the Principle of Personal Good

Option-dominance and the Principle of Personal Good 1

The Principle of Personal Good—that for all alternatives ϕ and ψ , if everyone is at least as well off under ϕ as under ψ , and someone is strictly better off, then ϕ is better than ψ —is the core of welfare economics. The PPG, however, is not well suited to situations in which individuals' responsibility for their welfare plays a role in assessing their claims. For the PPG is concerned with individuals' welfare outcomes, whereas once we take account of individuals' responsibility for their welfare, our attention shifts from their welfare outcomes to the quality of their options. It is then interesting to examine the relationship between responsibility-sensitive social choice rules and the PPG, to see whether the latter principle must be abandoned when we take account of individuals' responsibility for their welfare.

In this section, I will investigate the relation between the PPG and option-dominance. I examine the relation between the PPG and the Leximin Value of Option Set approach in section 4.8.

The contrast between the PPG and option-dominance becomes apparent when policies differentially affect the number of individuals who have chosen a certain level of education. Consider, for example, a policy that whilst keeping all rewards to

education the same as in the status quo, excludes Nationals from education beyond high school. Call this the ‘exclusion policy’ (see table 4.6). Imagine that in response to this policy, all Nationals who previously attended higher education now only complete high school. On the PPG, this exclusion policy is unambiguously worse than the status quo, since 83 million people are now worse off, and no one is better off. However, on option-dominance, the two policies are incomparable. This establishes that a policy ϕ can dominate a policy ψ by the PPG, but not option-dominate ψ .

Table 4.6 A policy that is PPG-dominated, but not option-dominated by the status quo.

	Exclusion policy					Total
	High School No Degree	High School w/Degree	College No Degree	Associate Degree	BA or more	
NATIONALS number ('000)	23,816	135,782	0	0	0	159,588
income	14,885	23,822	n/a	n/a	n/a	25,951
IMMIGRANT GROUP 1 number ('000)	4,655	7,581	4,812	1,401	3,398	21,847
income	11,948	19,934	24,445	29,115	42,361	25,306
IMMIGRANT GROUP 2 number ('000)	1,299	1,985	1,541	598	3,093	8,516
income	13,129	20,199	22,330	28,769	49,966	30,919
TOTAL number ('000)	29,770	145,148	6,353	1,999	6,491	189,951
income	14,349	23,232	23,932	29,001	46,469	24,429

It is also easy to see that a policy ϕ can option-dominate a policy ψ but not dominate ψ by the PPG. For any policy that, starting from the status quo, redistributes income from a relatively well off person at a given level of education to a less well off person at that same level of education, will option-dominate the status quo. (The ‘maximise value of option sets’ policy in table 4.3 is an example.) But since

this redistributive policy involves both gainers and losers, the PPG does not say that one is better than another.

We have therefore established that option-dominance and the PPG are logically independent of one another, since one principle may regard two policies as incomparable while the other expresses a strict preference between them.

Is it ever possible that the two principles conflict, in the sense that policy ϕ option-dominates ψ , while ψ dominates ϕ by the PPG? As it turns out, this is impossible: if ϕ option-dominates ψ , then ψ cannot dominate ϕ by the PPG. The proof of this result is provided below. It is not easy to provide a quick intuitive sketch of it, though the proof is followed by an illustration of how one proceeds in simple cases, which should give an idea of the overall strategy of the proof. One can get an initial feeling for the result by realising that for ϕ to option-dominate ψ , there must be at least one option, chosen by at least one person, for which the welfare under ϕ is larger than the welfare in ψ . If all individuals choose the same options under both policies, at least one person must therefore be better off under ϕ than under ψ . It is therefore clear that ψ cannot dominate ϕ by the PPG. The proof basically extends this simple case to cases where people choose different options under both policies.

It follows from these results that while a concern for the quality of individuals' options (as formalised by the criterion of option-dominance) does not require accepting the PPG, it also does not require abandoning the PPG.

* *Option-dominance and the Principle of Personal Good 2*

We claim that ψ does not PPG-dominate ϕ , if ϕ option-dominates ψ .

Proof: Suppose that ϕ option-dominates ψ , while ψ PPG-dominates ϕ in the strong sense, i.e. ψ is at least as good as ϕ for all individuals and better for some. For any individual n ($1 \leq n \leq N$), define

$w(n) := w(\phi, y_n, z_n)$ and $v(n) := w(\psi, y_n, z_n)$. We will show (*) that, for all

n ($1 \leq n \leq N$), there exist permutations σ, τ of $\{1, \dots, N\}$ such that (i)

$w \circ \sigma(1) \leq \dots \leq w \circ \sigma(N)$, (ii) $v \circ \tau(1) \leq \dots \leq v \circ \tau(N)$, and (iii)

$\sigma(k) = \tau(k)$ and $w \circ \sigma(k) = v \circ \sigma(k) = w \circ \sigma(n)$ for all $1 \leq k \leq n$. We first

show that the claim holds for $n=1$. Let σ, τ be permutations that satisfy (i)

and (ii). Let $j := \sigma(1)$ and $i := \tau(1)$. By option-dominance, $w(j) \geq v(i)$. By

PPG-dominance, we have $v(i) \geq w(i)$, and, by (i), $v(i) \geq w(j)$. Hence,

$v(i) = w(i) = w(j)$. Let σ' agree with σ except for interchanging the

position of i and j (i.e., except for $\sigma'(1) := i$ and $\sigma'(\sigma^{-1}(i)) := j$). σ'

and τ now satisfy conditions (i) through (iii).

Assuming that the claim holds for n , we will now establish it for

$n+1 \leq N$. Let $i := \tau(n+1)$ and $j := \sigma(n+1)$ and assume that $w(j) > v(i)$.

By PPG-dominance, $w(j) > v(i) \geq w(i)$ and, by (i), there exists some

$1 \leq k < n$ with $\sigma(k) = i$. By (iii), $\tau(k) = i$ and, since τ is a permutation,

we cannot have $i = \tau(n+1)$. By reductio, we conclude that $w(j) \leq v(i)$.

Option-dominance implies $w(j) \geq v(i)$. We thus obtain $w(j) = v(i)$ and,

by PPG-dominance and (i), $v(i) = w(i)$. Let σ' agree with σ except for

interchanging the position of i and j (i.e., except for $\sigma'(n+1) := i$ and $\sigma'(\sigma^{-1}(i)) := j$). σ' and τ now satisfy conditions (i) through (iii) of the claim. Having established claim (*), we note that it contradicts PPG-dominance (and, for that matter, option-dominance). Hence, our assumption is refuted and the proof completed. ■

We can illustrate this proof as follows. Consider first the case of one individual, \mathcal{A} , with one and the same option under both policy ϕ and policy ψ , so that he has the same responsible characteristic y_i under both. Suppose further that ψ PPG-dominates ϕ . In this case, it is obvious that ϕ cannot option-dominate ψ .

Consider next the case of one individual, \mathcal{A} , with two options under both policy ϕ and policy ψ , leading to responsible characteristics y_1 and y_2 . If \mathcal{A} chooses the same option under both policies, then PPG- and option-dominance cannot contradict each other. If \mathcal{A} chooses a different option under ϕ than under ψ , then the two are incomparable by option-dominance, so PPG- and option-dominance again cannot contradict each other. This establishes the claim for the case of one individual and two options. ■

Consider next the case of two individuals, \mathcal{A} and B who both have two options under both policy ϕ and policy ψ , leading to responsible characteristics y_1 and y_2 . There are now two relevant possible distributions of responsible characteristics under ϕ : (I) \mathcal{A} and B both choose y_i ; and (II) \mathcal{A} chooses y_1 and B chooses y_2 . (Whatever we prove for these two possibilities holds by analogy for the other two possible distributions of responsible characteristics.) I will discuss these cases in turn, assuming again that ψ PPG-dominates ϕ .

Case (I): Suppose (I-i) under ψ , A and B both have responsible characteristics y_1 . Then ϕ obviously cannot option-dominate ψ . Suppose (I-ii) that either A , or B , or both have responsible characteristic y_2 under ψ . Then ϕ and ψ are option-incomparable.

Case (II): Suppose (II-i) that A chooses y_1 and B chooses y_2 under ψ . Then ϕ obviously cannot option-dominate ψ . Suppose (II-ii) that both A and B choose y_1 or both A and B choose y_2 under ψ . Then ϕ and ψ are option-incomparable. Suppose (II-iii) that A chooses y_2 and B chooses y_1 under ψ . Then either (II-iii-a): A 's welfare under ϕ was equal to B 's welfare under ϕ , or (II-iii-b) A 's welfare under ϕ was greater than B 's welfare under ϕ , or (II-iii-c) A 's welfare under ϕ was less than B 's welfare under ϕ .

Now, in case (II-iii-a), given our assumption that ψ PPG-dominates ϕ , either A 's welfare under ψ is higher than under ϕ or B 's is, or both A 's and B 's welfare under ψ is higher than under ϕ . But this means that the least well off person at either y_1 or y_2 under ψ must be better off than under ϕ . So ϕ cannot option-dominate ψ .

In case (II-iii-b) again, either (II-iii-a-1) A 's welfare under ψ is higher than under ϕ and B 's is the same, or (II-iii-a-2) B 's is higher while A 's is the same, or (II-iii-a-3) both A 's and B 's welfare under ψ is higher than under ϕ . If (II-iii-a-1), then the least well off y_2 -individual under ψ is better off than under ϕ , so that ϕ cannot option-dominate ψ . If (II-iii-a-2), then the least well off y_2 -individual under ψ is better off than under ϕ , so that ϕ cannot option-dominate ψ . If (II-iii-a-3) then again, the least well off y_2 -individual under ψ is better off than under ϕ , so that ϕ

cannot option-dominate ψ . This completes the proof for the case of two individuals and two options. ■

4.7 Problems with the Leximin Value of Options view

As mentioned, the core idea of the Leximin Value of Options View is that we should give priority to those who are less well off than others who have made the same choices. By focusing on comparisons between individuals who have made the same choices, this approach picks out one way in which an individual may be disadvantaged by brute luck. But it also neglects a different way: an individual may be disadvantaged not because the welfare outcome of the option he chooses is less good than that of others, but because his option set is less valuable than that of others, either because (case (a)) he does not have good options that others do have, or because, (case (b)) while he has the same options that others have, the outcomes of these options are all relatively low. By way of illustration of case (a), consider again the ‘exclusion’ policy of table 4.6. Under this policy, Nationals are intuitively worse off than members of other ethnic groups due to no choice or fault of their own, not because they are worse off compared to members of these groups who make the same choices—indeed, they remain better off than members of these groups who either drop out of high school or complete high school only—but because they do not have good options that members of these other groups do have.

Now, the Leximin Value of Options view cannot register this way in which Nationals are disadvantaged. Instead, it sees Nationals who drop out of high school, and Nationals who complete high school only as being relatively advantaged. To see this, consider the ‘exclusion

plus transfer' policy depicted in table 4.7. This is the same as the previous exclusion policy, except that in addition, a small tax is levied on all Nationals and transferred in the form of a 1,000 unit grant to members of other ethnic groups. Now, by the criterion of option dominance, the 'exclusion plus transfer' policy will be better than the 'exclusion policy', because the worst-off at each educational level are better off than under the status quo. However, intuitively, the 'exclusion plus transfer' policy is worse than the 'exclusion policy', since it makes individuals who are already worst off due to brute luck (since they are deprived of good options that others have) even worse off.

Table 4.7 The exclusion plus transfer policy

	Exclusion plus transfer					
	High School No Degree	High School w/Degree	College No Degree	Associate Degree	BA or more	Total
NATIONALS						
number ('000)	23,816	135,782	0	0	0	159,588
income	14,647	23,584	n/a	n/a	n/a	22,249
inc. change*	-238	-238	n/a	n/a	n/a	-238
IMMIGRANT GROUP 1						
number ('000)	4,655	7,581	4,812	1,401	3,398	21,847
income	12,948	20,934	25,445	30,115	43,361	24,306
inc. change*	1,000	1,000	1,000	1,000	1,000	1,000
IMMIGRANT GROUP 2						
number ('000)	1,299	1,985	1,541	598	3,093	8,516
income	14,129	21,199	23,330	29,769	50,966	31,919
inc. change*	1,000	1,000	1,000	1,000	1,000	1,000
TOTAL						
number ('000)	29,770	145,148	6,353	1,999	6,491	189,951
income	14,509	23,431	24,932	30,011	47,469	24,429

* compared to 'exclusion policy'.

Let us now turn to case (b), in which individuals have the same options that others have, but the outcomes of these options are all relatively low. The pattern of rewards to education listed in table 4.8 gives a case of this kind. Nationals who are high school dropouts, who complete only high school, or who complete high school but drop out of college are much better off than members of both other immigrant

groups who make comparable choices, while members of Immigrant Group 2 are worst off at these education levels. At higher levels of education, the picture is different: members of Immigrant Group 2 are much better off than both Nationals and members of Immigrant Group 1, and Nationals are now worst off compared to others with similar levels of education. At each education level, members of Immigrant Group 1 have an income that is just slightly better than the income of the least well off individuals at that education level.

Table 4.8 A second problem case for the Leximin Value of Options View

	High School No Degree	High School w/Degree	College No Degree	Associate Degree	BA or more
NATIONALS income	25,000	30,000	35,000	<i>40,000</i>	<i>40,000</i>
IMMIGRANT GROUP 1 income	10,500	20,500	30,500	40,500	40,500
IMMIGRANT GROUP 2 income	<i>10,000</i>	<i>20,000</i>	<i>30,000</i>	50,000	60,000

Note: The italicised cells give the minima at each education level.

Again, intuitively, when we consider the value of individuals' option sets taken as a whole, it seems reasonable to consider members of Immigrant Group 1 as worst off. However, because the Leximin Value of Options view looks only at each education level separately, it registers only that members of this group are second least well off at each educational level. It would thus applaud a small transfer from each member of this group for the benefit of Nationals and members of Immigrant Group 2. However, from the perspective of the overall value of individuals' option sets, this would involve making those who are already worst off even worse off.

4.8 The Leximin Value of Option Set view

We have seen that a concern for brute luck disadvantage in the value of individuals' options will sometimes conflict with a concern for brute luck disadvantage in the quality of individuals' option sets. I believe that in cases where individuals can choose under adequate conditions of choice, eliminating the latter form of brute luck disadvantage is more important. For suppose we are in the situation characterised by the 'exclusion policy' of table 4.6, and consider the force of the possible responses to the two following complaints: First, consider the response of a supporter of the Leximin Value of Option Set view to the complaint of a member of Immigrant Group 1 who drops out of high school and says that he should be counted among those most disadvantaged by brute luck, because he is worse off than members of the other ethnic groups who made the same choice. To this complaint, one could respond: 'Though you are less well off than others who made the same choice, you had good and accessible options that some others did not have. If you had chosen these options, you would have been better off than those who were deprived of these options. So though the fact that you are worse off than others is in part due to brute luck, it is also due to your choice. Moreover, the obstacles you faced in becoming at least as well off as others were not greater than those faced by others. You are therefore not among those who are most disadvantaged by brute luck.'⁴⁸ This is a forceful response.

Consider next the response on behalf of a defender of the Leximin Value of Options view to the complaint of a National who says that he should be counted

⁴⁸ A similar response can be made to a member of Immigrant Group 2 who drops out of high school in the case characterised in table 4.8.

among those most disadvantaged by brute luck, because he is was deprived of good options that others had: ‘Though you did not have relatively good options that others did have, the value of the options you did have is greater than the value of others’ comparable options. You are therefore not among those who are most disadvantaged by brute luck.’ This response obviously lacks force.

In sum, when individuals can choose between their options under adequate conditions of choice, it is appropriate to refer to the quality of options that they did not choose, but could have chosen, in assessing their moral claims. This conclusion favours the Leximin Value of Option Set view over the Leximin Value of Options view. It is also, of course in line with our conclusion in chapter 3 that a form of egalitarianism that gives individuals’ choices and options a fundamental justificatory role should assess individuals’ claims by the value of their option sets.

The Leximin Value of Option Set view and the PPG

I now want to briefly remark on the implication for the PPG of accepting a moral theory that assesses individuals’ situation by the quality of their option sets.

It seems that a theory that is fundamentally concerned with the value of individuals’ option sets, rather than their welfare outcomes, will sometimes involve judgements that conflict with the PPG. For consider a case in which the value of person *A*’s option set is relatively low, because she only has options that yield relatively bad outcomes. Suppose we can greatly improve the value of person *A*’s option set by creating a good option for her by using resources gained by taxing individual *B*, who is relatively well-off due to brute luck. Let us suppose that this decreases *B*’s welfare, but does not make his option set less valuable than *A*’s will be after the improvement

in her option set. On the Leximin Value of Option Set view, we should tax *B* to offer *A* the relatively good option, even if we can foresee that person *A* will not choose this good option. This contradicts the PPG, which would require us not to tax *B* when this does not improve anyone's welfare. I conclude that if we want to give substantive responsibility a fundamental role, the PPG should be abandoned.

Conclusion

In sum, I have argued that when individuals can choose between the options in their option set under adequate conditions of choice, it is appropriate to evaluate their situation by the quality of their option set as a whole. A responsibility-sensitive egalitarianism that aims at improving the situation of individuals who are most disadvantaged by brute luck should therefore give priority to individuals with the least valuable option set.

Chapter 5. An Egalitarian Ethos?⁴⁹

“Rightly understood, these clauses [of the social contract] can all be reduced to one alone, namely, the total alienation of each associate with all his rights to the whole community (...). Furthermore, since the alienation is made without reservation, the union is as perfect as it can be, and no associate has anything more to claim. For if some rights were left to individuals, and there were no common superior who could decide between them and the public, each person, being in some respects his own judge, would soon claim to be so in every instance; the state of nature would subsist, and the association would necessarily become tyrannical or ineffectual. (...) It is agreed that each person alienates through the social pact only the part of his power, possessions, and liberty that is important to the community, but it must be agreed that the sovereign alone is the judge of what is important” (Rousseau 1988, pp. 92-93 and 101-102).

Introduction

In this chapter, I want to examine to what extent egalitarian justice requires that individuals use the opportunities they are provided with in particular ways. My

⁴⁹ Parts of this chapter will appear as ‘Incentives and Principles for Individuals in Rawls’s Theory of Justice’ in *Ethics and Economics*. Early versions of this chapter were presented at the Brave New World Conference on Political Theory at Manchester University in July 2001 and the Annual Meeting of the Dutch-Flemish Political Science Association in June 2002. I thank the participants in these meetings, and Brian Barry, Jerry Cohen, Peter Dietsch, Carina Fourie, Véronique Munoz-Dardé, Michael Otsuka, Robert van der Veen and Martin Wilkinson for their comments.

central question is: should citizens of a just egalitarian society pursue their egalitarian ideals through the design of equality-promoting formal public institutions only, or should they also sometimes be guided by egalitarian ideals in their individual choices within the space permitted for free individual choice by these formal public institutions?

The egalitarian ideals I refer to here are the two ideals we reviewed in chapter 2: a social and political ideal of equality, and a brute luck egalitarian ideal. I want to examine whether each of these ideals requires a choice-constraining ethos, and what the content of this ethos might be.

Scheffler (2003) suggests that while the social and political ideal of equality clearly has a place for an ethos of this kind, it is difficult to see what reason a brute luck egalitarian would have to advocate it. He writes:

“There is a (...) tension between the luck-egalitarian attitude towards choice and the attitudes associated with what Cohen calls an “egalitarian ethos”. The emphasis on the importance of a choice-constraining egalitarian ethos is quite congenial to the social and political ideal of equality, but the [brute] luck-egalitarian motivation for such an emphasis is less clear” (p. 37n77).

I believe Scheffler is mistaken on this point. It is, in fact, easy to imagine cases in which a brute luck egalitarian might want to introduce a choice-constraining duty of this kind. First, a choice-constraining ethos might be the best way to achieve brute luck equality. Imagine a society in which the single source of brute luck inequality is a genetic ailment which causes liver failure. If all members of this society live moderately healthy lifestyles, then upon their deaths, there will be just enough organs for transplantation to ensure that everyone with this ailment can be fully restored to health. If they do not, brute luck equality can only be achieved by cutting short the

lives of otherwise healthy individuals. Suppose further that because of limitations of public knowledge, a healthy lifestyle cannot be effectively enforced by the state. In this scenario, a brute luck egalitarian could endorse a choice-constraining duty for the purpose of achieving brute luck equality in an acceptable way.

Second, it is likely that the value of the least valuable option set, as well as the overall degree of inequality of the value of option sets, will be determined by individuals' choices from their option sets. A standard example is one in which in a society with relatively productive and relatively unproductive individuals maximises the value of the least valuable option set through a tax-and-transfer scheme. Under certain conditions, if the more productive individuals choose to work harder than they would if they made their labour/leisure decisions purely on the basis of their own interests, then this will generate greater tax revenues which can be used to improve the value of the least valuable option set, whilst simultaneously decreasing inequality in the value of individuals' option sets. A principle requiring choices of this kind might therefore be a requirement of brute luck egalitarian justice.

Indeed, it seems natural to assume that egalitarian justice should require both equality-promoting formal public institutions and an egalitarian ethos. For why should the pursuit of justice be limited to formal public institutions, and the behaviour required to support these institutions? As our two simple examples illustrate, it is easy to think of cases in which brute equality would be well served by the adoption of an egalitarian ethos. Similarly, it is easy to think of cases in which the maintenance of egalitarian social relations may require constraints on behaviour, like treating others with respect in one's everyday interactions, that are not required by these institutions.

Nonetheless, recent debates on the scope of justice in Rawls's work have thrown up some challenges to this view.⁵⁰ In this chapter, I will review some of this debate in order to examine what we can learn from it about the roles of formal public institutions and principles for individual conduct in egalitarian justice.

In section 5.1, I review Cohen's criticism of Rawls's limitation of the scope of his three principles of justice, including the difference principle—the principle regulating the distribution of income, wealth, and the social basis of self-respect—to the major legal and socio-economic institutions of society, or what he calls its “basic structure”.

In section 5.2, I discuss Philippe van Parijs's (2003, pp. 226-231) defence of Rawls's restriction of the scope of the difference principle on the grounds that it follows from a general restriction of the scope of justice to principles for the basic structure, and that this general restriction, in turn, follows from the demands of political liberalism. In sections 5.3 and 5.4 I challenge Van Parijs's interpretation and defence of Rawls by showing that Rawls's contractualism does not restrict the scope of principles of justice to the basic structure, and that certain principles for individual conduct that go beyond individuals' duties to establish and support just institutions are an integral part of Rawls's political conception of justice.

In section 5.5, I consider Rawls's own reasons for not supplementing the principles for individual conduct that are part of his theory with further principles for individual conduct, such as a duty to take account of the impact of one's economic

⁵⁰ See Cohen (1995, 1997, 2000, 2003, 2005). For some responses to Cohen's criticism, see Andrew Williams (1998), David Estlund (1998), Thomas Pogge (2000), Frank Vandebroucke (2001), Saladin Meckled-Garcia (2002) and Norman Daniels (2003). Cohen's criticism represents the revival of an earlier debate. For a review of this debate and a critical contribution, see Barry (1989, appendix C).

choices on the most disadvantaged, or to promote an egalitarian and democratic culture in one's everyday actions. I argue that Rawls's theory does not contain such further duties for two reasons. First, a concern with the publicity of the principles of justice: that people know what they demand of themselves and of others, and can know that most people comply with their demands. Second, a concern for basic liberties like freedom of association, or freedom of occupational choice.

In response to Rawls's arguments, I argue that though publicity identifies an important constraint on the formulation of principles for individual conduct, it does not rule out adding further principles to the ones Rawls accepts. I also argue that principles for individual conduct do not limit liberty.

Since the arguments from publicity and liberty fail, and in light of the beneficial consequences of the adoption of such an ethos, I conclude that an egalitarian ethos encompassing both the principles which Rawls identifies, and further principles regulating individual conduct and motivation should be seen as part of an egalitarian theory of justice.

5.1 The scope of the difference principle and Cohen's critique

Against a background of equal basic liberties and fair opportunity for education and jobs, Rawls proposes that in a just society, socio-economic inequalities should be regulated by the difference principle: inequalities are just if and only if they are necessary to make the worst-off as well off as they can be.⁵¹ In several passages

⁵¹ See TJ (p. 72, emphasis added): "Social and economic inequalities are to be arranged so that they are both (a) to the *greatest benefit of the least advantaged* and (b) attached to offices and positions open to all

Rawls writes that the scope of the difference principle is confined to what he calls the “basic structure of society”. Rawls defines the basic structure as “a public system of rules defining a scheme of activities that leads men to act together so as to produce a greater sum of benefits and assigns to each certain recognised claims to a share in the proceeds” (TJ, p. 74) and also as “the way in which the major social institutions fit together into one system, and how they define fundamental rights and duties and shape the division of advantages that arises through social cooperation” (PL, p. 258).

Now, a public system of rules in this sense can, of course, include rules for individual conduct that are not formally laid down in laws or enforceable rules, as long as these rules of conduct are public, that is, can be understood by everyone and can be known to be generally followed. But, Rawls’s more concrete specifications of the basic structure do not mention such informal rules. Instead, as examples of institutions that belong to the basic structure, he offers “the political constitution, the legally recognised forms of property, the organisation of the economy, and the nature of the family” (PL, p. 258). Aside from the case of the family, these examples all suggest that Rawls confines the basic structure to what Frank Vandembroucke (2001, chapter 7) calls *formal* institutions: spoken, written or tacitly understood statements that carry firstly, a known range of sanctions, second, a norm or rule that prescribes these sanctions, and thirdly, provisions for monitoring, all of which emanate from a public rule-making arena like a court or government. Furthermore, Rawls’s discussions of the sense in which the family is part of the basic structure suggests

under conditions of fair equality of opportunity.” For identical formulations, see also TJ (p. 245), PL (pp. 6-7 and 291) and CP (p. 258). Van Parijs (2003) offers an extended discussion of different formulations of the difference principle.

that in this case, too, he has in mind only the formal regulation of the institution of the family (JF, pp. 162-168; CP, pp. 595-601).

Taking Rawls's understanding of the basic structure to be limited to the major social formal institutions allows us to make a clear distinction between the institutions of the basic structure on the one hand, and the possibly but not necessarily public, informal rules and strategies that individuals and voluntary associations use to regulate their conduct within the limits prescribed by the institutions of the basic structure. We then get a relatively clear idea of the scope of the applicability of Rawls's three principles of justice, including the difference principle: they apply to the major social formal institutions, and not to the actions of individuals and associations within the limits prescribed by the institutions of the basic structure. As Rawls puts it:

“Thus, the principles of justice, in particular the difference principle, apply to the main public principles and policies that regulate social and economic inequalities. They are used to adjust the system of entitlements and earnings and to balance the familiar everyday standards and precepts which this system employs. The difference principle holds, for example, for fiscal and economic policy. It applies to the announced system of public law and statutes and not to particular transactions or distributions, nor to the decisions of individuals and associations, but rather to the institutional background against which these transactions and decisions take place.” (PL, pp. 282-283).⁵²

Because individuals are left free to pursue their permissible conception of the good within the rules specified by the basic structure, the institutions that must

⁵² See also PL (p. 284) and JF (p. 73).

conform to the difference principle will offer incentives to elicit from people, pursuing their own ends, those kinds of choices that will enhance the position of the worst off. These incentives will result in inequalities, but these are just, according to Rawls, since institutions have been set up to maximise the level of social and economic benefits the least advantaged can expect, consistent with the basic liberties and fair opportunity being established to a sufficient degree.⁵³

Cohen challenges the coherence of this separation of institutions from personal behaviour as follows. If the difference principle countenances only *necessary* inequalities, then many incentive-based inequalities are not justified. For if people were motivated to base their choices in the economic and personal sphere on their impact on the worst-off, then they would have little need for inequality-producing incentives. Little need, for, Cohen (2000, p. 206n24) writes, it is reasonable to assume some scope for purely personal projects. (It is important to keep in mind that these projects need not be selfish—they may involve the fulfilment of religious obligations, for example, or the care for friends or family members.) But this limited scope is very different from the unlimited space for the pursuit of one's goals that Rawls permits. The difference principle, Cohen concludes, requires a personal ethos which, with allowance for some personal prerogative, leads people to base their economic decisions about how much to work and which career to pursue (and perhaps other

⁵³ See, for example, TJ (p. 68), JF (pp. 63-64). I include the phrase 'to a sufficient degree', since I agree with Van Parijs (2003) that a literal interpretation of the lexical priority of establishing a fully adequate scheme of basic liberties and establishing fair opportunity is implausible. Strict lexical priority would entail there would be no resources left to the distribution of which the difference principle would apply. For it is always possible to devote more resources to securing the basic liberties or fair equality of opportunity. Rawls recognises the approximate nature of the lexical ordering (TJ, p. 55). I return to this point in section 5.5.

personal decisions as well, such as how to treat one's children, whom to befriend and marry, etc.) on how these decisions affect the least well-off.

Cohen presents three arguments why this extended interpretation of the scope of the difference principle is the only one consistent with Rawls' full theory of justice. First, Rawls presents a just society as one in which people affirm and uphold the principles of justice. In the political sphere people are required not to further their personal projects, but to support those policies that meet the requirements of the principles of justice.⁵⁴ But a split personality would then seem to result between citizens as political actors and as actors in the economic and personal sphere (Cohen 2000, pp. 124-125).

Second, Cohen claims that the inequalities caused by acting on the basis of unconstrained pursuit of one's own projects in the economic, personal and associational spheres mean that three features of what Rawls sees as a just, or well-ordered society cannot be maintained (2000, pp. 134-136). The first of these features is fraternity, which is taken to mean that citizens only wish to enjoy greater benefits when this is to the benefit of all their fellows. For Rawls, this fraternity is realised in a society which is regulated by the difference principle:

“The difference principle, however, does seem to correspond to a natural meaning of fraternity: namely, to the idea of not wanting to have greater advantages unless this is to the benefit of others who are less well off. (...)

Now *wanting to act on the difference principle* has exactly this consequence. Those

⁵⁴ For this duty of citizens with regard to voting, see TJ (pp. 294 and 314). This duty follows from what Rawls calls 'the natural duty of justice', which requires individuals to support and comply with existing just institutions and to further the establishment of just institutions. See TJ (sections 19 and 51).

better circumstanced are willing to have their greater advantages only under a scheme in which this works to the benefit of the less fortunate.” (TJ, pp. 90-91; emphasis added).

Cohen takes the emphasised words to indicate that Rawls should believe such fraternity is only realised when the resulting distribution is not only a product of the way in which the basic structure is organised, but also of people’s attitudes and choices within this structure. Fraternity is thus not compatible with unlimited pursuit of personal goals in the economic and personal sphere.

The second feature is the fact that those that are less well off can bear their position with dignity. Cohen (2000, p. 135) claims that the awareness of the least well off that their predicament results from the fact that those better situated cannot be bothered to take their interests into account in their decisions means that their ability to bear their situation with dignity is undermined.

The final feature is the form of social union that Rawls believes a society governed by the principles of justice makes possible. In a “well-ordered” society, (nearly) everyone accepts and knows that others accept the principles of justice, and the basic social institutions satisfy and are known to satisfy these principles (TJ, pp. 4-5 and 397ff). The members of such a society “share the goal of giving one another justice”: they have a generally effective desire to act as the principles of justice require, and to adjudicate and justify their claims on social goods and on each other from the shared point of view of these principles.⁵⁵ Now, in such a society, Rawls writes, people enjoy a special form of social union: its citizens share a final end—to

⁵⁵ See JF (p. 20) for the shared end of giving one another justice. For the adjudication of competing claims, see TJ (pp. 4-5 and 397ff), CP (p. 250) and JF (p. 9). For the interpersonal justifiability of claims, see TJ (pp. 145, 297 and 510).

set up institutions that accord with the principles of justice and to act as these principles require—and they see this end as good in itself, because it enables them to achieve the full realisation of their nature as social and moral beings. In addition, a society governed by the principles of justice allows people to pursue their diverse aims as individuals and within voluntary associations against a background which ensures that they have an acceptable amount of the goods necessary to develop their capacities. Because Rawls takes it as a fact of human psychology that we take pleasure in the development and exercise of diverse human capacities when this takes place in a just society, such a society must be experienced as good (TJ, pp. 462-463 and JF, pp. 198-202).

Now, the central prerequisite for the realisation of this form of social unity is that all members of society are willing and able to justify their institutions, attitudes and behaviour to each other from a common perspective. Cohen, however, claims that a society in which people, within the space permitted by the rules of the basic structure, are allowed to pursue their own projects without limitation must fail this test of interpersonal justifiability, since those placed in a favourable position cannot justify to the worse-off members of society demanding inequality-producing incentives in order to make socially beneficial choices to invest in particular ways and choose and perform well in useful jobs, etc. (1995, p. 350; 2000, p. 135). Cohen's second argument concludes that anyone who wishes to maintain that the difference principle does not apply to people's motivation and behaviour generally must give up these three features of a Rawlsian well-ordered society.

The third argument questions the foundation of what he sees as Rawls's idea that justice applies only to major social institutions, and not to behaviour within these institutions. As we saw above, Rawls appears to limit the basic structure to the major

social formal institutions, or, as Cohen calls it, to the “broad coercive outline of society”. However, such limitation of what is meant by the basic structure falls afoul of Rawls’s criterion for focusing on institutions: the profound effects of these institutions on people’s life chances (Cohen 2000, p. 138 and TJ, p. 7). By this criterion, we cannot confine ourselves to the major social formal institutions, since many institutions and patterns of individual choice that fall outside of the basic structure so defined, such as role patterns in the division of labour and practices of favouring the education of children of one sex or level of ability over another obviously do have profound effects on the distribution of benefits and burdens in society. So, Cohen concludes, we cannot coherently maintain that people’s choices and motivation are not as much the subject of justice as the major social institutions (2000, pp. 131-136; 2003, pp. 91-92).

Of these three objections, Cohen considers the first ‘split-personality’ one only in passing, and he believes the second means only that a Rawlsian must abandon any homilies about fraternity, equal dignity and the realisation of social unity as so much icing on the cake of a well-ordered society, not that the project of separating institutions from personal choices must fail. It is the third objection that he sees as decisive against any conception of justice that excludes any element of people’s behaviour from its purview (2000, pp. 130 and 136).

I will focus on possible responses to this critique in sections 5.2 and 5.5. Before proceeding, a remark on fraternity and the dignity of the worst off. In the case of fraternity, we must note first that unlike equality and liberty, the incorporation and interpretation of the idea of fraternity within the theory of justice is not central to its justification. It is, rather, offered as one advantage among many of the difference principle that it *seems* to offer an interpretation of the concept and the principles

required to apply it. In addition, it is a rather idiosyncratic interpretation of the idea of fraternity that Rawls claims the difference principle represents. Consistent with Rawls's general scepticism about the role that affective sentiments can play in sustaining individuals' motivation to act justly,⁵⁶ his interpretation of fraternity seems not to demand the affective ties which are generally believed to be part of the idea (TJ, pp. 90-91). Rather, it appears to apply first and foremost to "institutions and policies", which must satisfy its demands. This sense of fraternity can be realised in a society in which people do not recognise a duty of the kind Cohen advocates.

Rawls's remarks on dignity have to do with the moral worth of the worst off. In the context of discussing the idea that distributive justice might be taken to mean that people should get what they morally deserve (where moral desert is understood as a concept that has application prior to distributive institutions that follow from of a conception of justice), Rawls remarks that the inequalities sanctioned by the difference principle have nothing to do with such pre-institutional moral desert (TJ, pp. 273-277). Thus, Rawls's point is merely that the dignity of the worst-off need not be undermined by inequalities that result from the application of the principles of justice, because these do not express unequal moral worth.

5.2 Van Parijs's defence of Rawls

Van Parijs (2003, pp. 226-231) has argued as follows that Rawls's restriction of the scope of the difference principle reflects the demands of political liberalism. In *A Theory of Justice*, Rawls presents a "complete conception of right", which is a

⁵⁶ See, among others, TJ (pp. 112 and 155), PL (p. 87) and JF (p. 182).

conjunction of “justice as fairness” applying to the institutions of the basic structure and “rightness as fairness” applying to individuals’ actions within the space permitted by these institutions. It would be natural to think that this complete conception of right must contain a principle requiring individuals to make their economic choices in ways that would improve the lot of the worst off, given the potential of such a principle to improve the situation of the worst-off and decrease inequality. However, in *Political Liberalism* (p. xlii) and *Justice as Fairness* (p. xvii), Rawls argues that political justice should not be seen as a complete conception of right, and instead should be concerned only with demands of justice as fairness. This means that the Rawls of *Political Liberalism* must confine himself to principles for the institutions of the basic structure. For requiring certain principles for individual conduct beyond those needed to support and maintain just institutions would be to cross the line from a political conception of justice—one that is independent of and without grounding in premises peculiar to metaphysical, epistemological, and general moral conceptions, and which can be shared by people who hold very different conceptions of these kinds—into a comprehensive moral conception.

Van Parijs worries that this means that the demands of political liberalism leave us with “exceedingly feeble redistributionary mechanisms” (2003, p. 230). But he also discerns a solution to this problem in the way institutions can affect people’s motivation and behaviour:

“social policies, labour market legislation, and the regulation of credit and advertising may conceivably encourage or discourage, to very different extents, an ethos of work and thrift. (...) Promoting the work ethos of the more skilled and affluent [is good for the worst off], as part of their greater

output, as opposed to their leisure, can be used to boost the expectations of the worst off? (2003, p. 231).

Thus, Van Parijs believes, without crossing the boundary between a purely political conception of justice and a comprehensive moral doctrine—a boundary marked by the distinction between principles for the basic structure of society and principles for individual conduct within the space permitted by the basic structure—the difference principle can be saved from being enfeebled by self-seeking individual conduct by “resolutely designing institutions that foster an ethos of solidarity, of work” (2003, p. 231).

In sections 5.3 and 5.4 I will argue that Van Parijs’s idea that the ‘political liberalism/comprehensive moral doctrine’ divide is marked by the ‘principles for the basic structure/principles for individual conduct’ divide is mistaken. I do so by showing that certain principles for individual conduct that go beyond individuals’ duties to establish and support just institutions are an integral part of Rawls’s political conception of justice.

5.3 Total Alienation

In the opening lines of *On Social Contract*, Rousseau explains that he wishes to determine the principles of legitimate social institutions, starting from facts about human nature and the kinds of institutions to which human beings can conform (1998, p. 88). In the section quoted at the start of this chapter, Rousseau announces that in considering these principles, we must start from a position of “total alienation

by all”. I shall argue that Rawls’s contractualist enterprise also demands this “total alienation”. But what does the phrase mean?

Rawls’s social contract view at the outset delimits the reasons for which we adopt principles of justice for institutions and individuals. For the contractualist, justice has a particular role. Society is marked by identity and conflicts of interests between its citizens. There is identity of interests because social co-operation makes possible for everyone a life that is much better, both in a material and moral sense, than a life outside of society. There is a conflict of interests because, abstracting from moral considerations, people all prefer a larger to a smaller share of the benefits of social co-operation, and because they have different moral, religious, and ideological views. The role of a theory of justice is to ensure that in spite of this conflict, effective and just social co-operation can come about. This does not merely mean the absence of open conflict, but the establishment of a community in which the distribution of benefits and burdens and the exercise of power take place in accordance with rules that free, equal and reasonable citizens would accept if placed in an initial situation of equality (TJ, section 1; PL, lecture 1, and CP, pp. 560-561). Rawls stresses the requirement that the general adherence to these rules can be judged. The possibility of such acceptance and judgement establishes a community of justification, necessary for the realisation of people’s moral nature. Because of the assumption that people have different moral views, such justification cannot be based on showing how the rules and institutions of society serve some communal end or ends. Society as a whole has no ends or ordering of ends, and principles of political justice can therefore not prescribe any ends to individuals or voluntary associations, except the end of giving others justice (TJ, p. 7; PL, p. 276; JF, pp. 10-12 and 20).

The social contract view, then, focuses exclusively on individuals engaged in social co-operation and subject to the exercise of coercive power by the state. These individuals are born into a society whose basic institutions and culture profoundly affect the development of their character and abilities, and from which it is difficult to exit. Rawls holds that these facts make a purely voluntary acceptance of existing social arrangements of society insufficient grounds for their legitimacy, which is one reason why the hypothetical social contract approach is necessary (PL, pp. 271-278).

Because of this focus, Rawls's theory of social justice does not pronounce on the justice or injustice of situations in which individuals are not involved in any co-operation, such as people living in complete isolation from each other. Does this focus also exclude from the scope of social justice principles which govern the activities of individuals and voluntary associations that do not share the three key characteristics of the political realm (its unchosen profound influence, its coerciveness, and lack of a shared ordering of ends)? Particular relationships of individuals, such of those of love and friendship, and particular associations, such as sports clubs and churches, are entered into and can be left voluntarily. Any coercive enforcement of agreements between individuals and within associations operates through the state, either directly or by the state giving the rights to such enforcement to a person or organisation. Lastly, individuals and associations have definite ends, the content of which can determine which principles of morality and "local justice" apply to them.

It would be a mistake, however, to think that these differences between the major social coercive institutions and these individual relationships and voluntary associations mean that the latter mark out separate spheres of life beyond the purview of social justice. For though the ends and ideals that shape these institutions

and relationships will be distinct from the end of social arrangements taken as a whole, the shape they take may profoundly influence the development of individuals' sense of justice—their ability and willingness to understand, propose and honour fair terms of cooperation if others will likewise do so—and the conditions under which they can form and pursue their conception of the good life. The kinds of relationships people can engage in, and the voluntary associations there are or can be in society are important determinants of the range of possibilities open to people to realise their plan of life.⁵⁷

Furthermore, the conditions under which these relationships and associations can take place are determined by society's coercive structure and the concomitant distribution of resources (TJ, pp. 93; 96-97; PL, p. 266). These have to ensure that the conditions under which people enter into their relationships and participate in associations do not undermine people's status as free and equal citizens, or their willingness and ability to support social justice. In addition, these rules have to ensure that the conditions under which they can join and leave are generally conducive to people's pursuit of their projects. The establishment of social justice therefore has priority over any principles that might apply to the non-political relationships of individuals and to associations (PL, p. 261, JF, pp. 10-12, 40, 162-168 and 182-183). These principles are bound by the restrictions specified by the principles of social justice. Of course, this does not mean that the principles of justice must dictate specific ends to these relationships and associations. For it is of great value to individuals that, within certain limits, individuals and associations are free to form and pursue their own ends (CP, p. 598).

⁵⁷ In this sense, the civic and cultural structure of a society functions like the other primary goods, and should perhaps be added to Rawls's list. See also the discussion in Kymlicka (2002, chapter 6).

In sum, any space in which principles specific to individuals and associations that are not part of social justice apply is the product of a prior establishment of the demands of social justice. As Rawls writes:

“A domain so-called, or a sphere of life, is not, then, something already given apart from the principles of justice. A domain (...) is simply the result, or upshot, of how the principles of political justice are applied, directly to the basic structure and indirectly to the associations within it. The principles defining the equal basic liberties and fair opportunities of citizens always hold in and through the so-called domains. (...) So the spheres of the political and the public, and of the not-political and the private, take their shape from the content and application of the conception of justice and its principles. If the so-called private sphere is a space alleged to be exempt from justice, then there is no such thing” (JF, p. 166; CP, p. 599).

We are now in a position to see what Rousseau’s remark about the “total alienation of each associate with all his rights to the community” means, and how it applies to Rawls’s contractualism. It means that though the principles of social justice are pre-contractually limited to the kind of role they are to play, they apply to all aspects of individuals’ lives and of social institutions that are relevant for establishing and maintaining just social co-operation. Insofar as we are concerned with establishing social justice, no aspects of people’s lives are pre-contractually excluded from falling under the scope of the principles to which everyone would agree in the initial contractualist situation of equality. Of course, at the level of application in social life certain parts of people’s personal and associational life may not be regulated by the principles of justice, but this space where the principles of social justice find no practical application is merely the upshot of a decision procedure

which at the contractual level takes no aspect of human society to be beyond the scope of social justice.⁵⁸

Since this space is determined by the contracting parties, it does not preclude the establishment of justificatory community. If, for example, the contracting parties would agree for certain reasons that the unlimited pursuit of one's own projects is permissible within the rules specified by the basic structure, then individuals could justify the pursuit of their projects within these rules to each other with reference to this agreement. (We will discuss the reasons Rawls believes contracting parties would have to allow such space in section 5.5). For the same reason, citizens will not need to develop a 'split personality' to cope with the different standards of the public sphere on the one hand, and the economic, private, and associational spheres on the other. For insofar as the demands of political justice are concerned, their actions in each sphere can be justified from a single perspective.

⁵⁸ Thus, the contractualist theory of Rousseau and Rawls contains two distinct 'levels': the contractual level, at which no one has any pre-established claim to any of the benefits and burdens of social co-operation, and the post-contractual society governed by the principles of justice in which individuals have such claims. It seems to me that interpretations of Rousseau's *On Social Contract* as totalitarian might in part be the result of a failure to distinguish these two levels. For the "total alienation" in the quoted passage does not imply that Rousseau's society will be totalitarian. The contracting parties, who must agree to principles that apply equally to all, will naturally want certain rights against undue interference in their lives, political rights and rights to (some) private property. These rights are discussed in chapter 4 of book II of *On Social Contract*. See also chapter 8 and 9 of book I and Gourevitch (1997, pp. xix-xx).

5.4 Principles for Individuals in Rawls's Theory

Throughout his work, Rawls refers to the basic structure as the “primary”, “first”, or “initial” subject of social justice, but never writes that it is the sole subject of social justice (TJ, pp. 6 and 47; PL, pp. 11-12 and 257; JF, p. 10). Indeed, Rawls writes that a suitable specification of the principles that apply to the basic structure and to the separate and free transactions between individuals and associations is essential to a theory of justice (JF, pp. 53-54; TJ, pp. 93ff and 293ff). Nevertheless, the initial focus in Rawls's theory is on the basic structure. Rawls gives several reasons for this focus, which I will only mention here.⁵⁹ The first set of reasons has to do with the basic structure's importance in shaping individuals' expectations and preferences. Taken together, the institutions of the basic structure are the most important determinants of the distribution of rights, obligations, benefits and burdens of social co-operation (TJ, p. 7). In addition, individuals' beliefs, attitudes and aims are shaped by the basic structure (TJ, p. 229; PL, p. 269). Second, Rawls argues that principles for individual behaviour alone cannot guarantee the requisite fairness in distribution, because principles for individuals alone would be epistemically too demanding to follow and too difficult to judge in their overall consequences (TJ, pp. 73-78; PL, p. 265 ff; JF, p. 54). Third, he indicates that all-encompassing principles for individual behaviour might undermine individuals' freedom to pursue their own conception of the good, while arranging the basic institutions of society to take care of background justice and then leaving individuals free to pursue their ends would preserve this freedom.⁶⁰ Finally, Rawls argues that since many principles for individuals presuppose some

⁵⁹ For a detailed discussion, see Vandenbroucke (2001, pp. 239-246).

⁶⁰ See Vandenbroucke (2001, pp. 240-241).

institutional structure to be in place, it is easier to discuss principles of social justice for individuals after those for institutions have been chosen (TJ, pp. 9-10 and 93-95).

Four sections of *A Theory of Justice* (18, 19, 51 and 52) are devoted to discussing principles for individuals. Rawls begins the first of these sections with the statement that “certain principles of this type [for individuals] are an essential part of any theory of justice” (TJ, p. 93). The difficulty is, however, that Rawls discusses these principles in the context of a complete conception of right, which he calls “rightness as fairness”, which he later describes as outside the bounds of a purely political conception of justice (TJ, 15 and 95-96; PL, xlii; JF, xvii). Does this mean that these principles for individuals have no place in Rawls’s purely political conception of justice, as Van Parijs believes? I do not think so. For the principles for individuals Rawls discusses in *A Theory of Justice* at any length all belong to the domain of social justice.⁶¹ Rawls considers certain principles for individuals part of social justice because these principles play an essential role in ensuring the stability of a just society. As Rawls writes in *A Theory of Justice* when he introduces his lengthier discussion of principles for individuals:

“I now wish to take up the principles of natural duty and obligation that apply to individuals. The first two sections examine the reasons why these principles would be chosen in the original position and their role in making social co-operation stable” (TJ 293).

⁶¹ For example, in *A Theory of Justice*, after mentioning the principle of fairness and the natural duty to support just institutions, Rawls writes: “I shall say very little about the other kind of principles for individuals. For (...) I must limit myself to the theory of social justice” (TJ, p. 100). See also JF (p. xvii): “(...) the problems examined in *Theory* in any detail are always the traditional and familiar ones of political and social justice”.

A society ordered by the principles of justice must be stable—must reproduce its conditions of existence, both materially and in its institutions and the attitudes and behaviour of its citizens—because it would be irrational for the contracting parties to agree to principles that could not be adhered to, since these would not be able to play the role justice is designed to play, i.e. ensure effective and fair co-operation in society.⁶² Therefore, if they contribute to the establishment of social justice, the principles for individuals developed in *A Theory of Justice* must survive the shift to a purely political conception of justice in Rawls's later work. Let us therefore look at the four principles for individuals which Rawls discusses in most detail, and trace their relation to the problem of stability.

Rawls divides the principles for individuals into what he calls “natural duties” and “obligations”. The former hold independently of voluntary acts, whereas the latter depend on a voluntary act to take part in an institution or enter into a particular relationship with individuals (TJ, pp. 96-97). I shall first look at three natural duties: the duty to support and further just institutions and the duties of mutual respect and mutual aid. I shall then look at what Rawls calls the “principle of fairness”, and one of the obligations that derives from it: to keep a bona fide promise.

The duty to support and further just institutions demands that individuals comply with the rules of just institutions, and aid in their establishment when this is not too costly. This includes the duty to vote and encourage others to vote for those parties and policies which best conform to the principles of justice (TJ, pp. 293-294). This duty is a natural complement to seeing justice as confined to the basic structure,

⁶² “They [the parties in the original position] are rational in the sense that they will not enter into agreements they know they cannot keep, or can do so only with great difficulty.” (TJ, pp. 125-126. See also section 29)

since it demands of us that we comply with the demands of the institutions of the basic structure, and that we make efforts to establish and maintain a just basic structure. But the duty asks nothing more of us, and is therefore consistent with the interpretation of Rawls that I am disputing.

However, two further duties that Rawls discusses at length cannot be seen in this light. The first is the duty of mutual respect (TJ, pp. 155-156 and 297). This duty demands that individuals recognise and treat one another as beings with a sense of justice and a conception of the good. It implies a willingness to see things from another's point of view, and to give reasons to others whenever their interests are materially affected, thereby implying that citizens should form a community of justification. It disallows contempt or lack of esteem for others, and asks that citizens treat each other courteously, and are prepared to do each other small favours. The second is the duty of mutual aid, which means that we are to help others who are in need when doing so is not too costly for us (TJ, pp. 100 and 297-298).

Now, these two duties are not just important because of the balance of good over bad that would result from their being generally followed. Rather, they are part of the principles of social justice because of the way they affect the relations between individuals and individuals' self-respect:

“Once we try to picture the life of a society in which no one had the slightest desire to act on these duties, we see that it would express an indifference, if not disdain for human beings that would make a sense of our own worth impossible” (TJ, p. 298).

It is this connection with self-respect that makes these principles part of a theory of social justice. Rawls writes that parties in the original position would want to avoid at almost any cost the conditions that undermine self-respect (TJ, p. 386). One

reason for this importance is that without a secure sense of self-worth, people's lives become meaningless to them: they can see no point in planning and managing their lives, or in undertaking any activity. A second reason is the role of people's sense of self-worth in establishing effective social co-operation, and it is this reason that concerns us here. Self-respect affects stability in two ways (TJ, p. 155). The first is that a society in which individuals support each other's sense of self-worth is a prerequisite for the development of citizens' sense of justice. Rawls's view of how citizens acquire an effective sense of justice depends on a three-stage theory of moral development (TJ, pp. 429-430). At the first stage, children raised in a caring environment form affective ties with the members of their family. At the second stage, finding that institutions are just and that people generally and with evident intention act on principles of justice, people develop trust in these institutions and their fellow citizens. At the third stage, realising that their interests and the interests of the people they care about are furthered by the arrangements of a just society, and their worth is affirmed, people acquire a normally effective sense of justice. Now, the duties we are discussing impact either directly, or indirectly through their influence on the sense of self-worth, at all three stages of this development of the sense of justice. Firstly, Rawls argues, people with a secure sense of self-worth are more likely to care for their children (TJ, p. 436). Secondly, other people's willingness to take our needs into account and treat us with respect contributes to the process of coming to trust them (the second stage) and to the fact that social arrangements are such that they further our good (the third stage).

The second way that self-respect influences the stability of society is through its relation to envy. Envy in the sense in which it involves ill will towards the person

envied⁶³ is what Rawls calls a “disruptive attitude” because it is collectively disadvantageous (TJ, p. 125). It leads the worse-off to wish to deprive the better off of their greater benefits, even at some cost to themselves. In turn, this leads the better off to take precautions against the hostile acts to which the worse-off become prone. Conscious of others’ negative attitudes towards their good fortune, they may even become spiteful: they become willing to give of goods of their own in order to deny the less well off certain benefits. People moved by envy and spite, then, will not have the motivation to adhere to fair and mutually beneficial arrangements.

Envy may arise in a well-ordered society because the social and economic inequalities sanctioned by the difference principle may be large (TJ, p. 446). If they are large enough, and in addition other social arrangements and individual behaviour serve to undermine the sense of self-respect of the worst-off, the circumstances in which society places these individuals can be such that it is unreasonable to expect them to overcome their rancorous feelings. Indeed, under such circumstances, the non-moral feeling of envy may give rise to the moral feeling of resentment, since people may resent being made envious (TJ, p. 468).

It is essential, then, that in a well-ordered society the conditions for strong feelings of the destructive form of envy do not arise. Rawls offers two reasons that they indeed will not. Firstly, he argues that in a competitive economy under conditions of fair opportunity, and subject to forms of taxation demanded by the difference principle, large differences in remuneration for scarce talents will not be permanent (TJ, pp. 136-137; JF, pp. 66-67). Secondly, he sees the main psychological source of envy as a lack of individuals’ confidence in their own worth:

⁶³ For the distinction between ‘benign envy’, which entails wishing we were in another’s situation without begrudging him his advantages, and ‘disruptive envy’, see TJ (pp. 466-467).

“someone sure of the worth of his plan of life and his ability to carry it out is not given to rancor nor is he jealous of his good fortune. Even if he could, he has no desire to level down the advantages of others at some expense to himself. This hypothesis implies that the least favored tend to be more envious of the better situation of the more favored the less secure their self-respect” (TJ, p. 469).

I will return to the first claim in section 5.5, and here remark only that the contribution of the recognition and adherence to the demands of the duties of mutual aid and mutual respect to citizens’ sense of self-worth will tend to lessen the tendency for envy to develop in society.

It seems natural to think that the duties of mutual aid and respect are not restricted to people’s actions within the institutions of the basic structure or their activities that pertain directly to the establishment and maintenance of a just basic structure. For one of their functions, to sustain mutual trust and confidence in one another’s intentions, cannot be fulfilled if individuals apply them so selectively. Not to have or act on racist attitudes, for example, follows from the duty of mutual respect (CP, p. 461). It would seem both very strange and in contradiction with Rawls’s stated purpose for taking this to be a duty to think that this requirement applies only when people are fulfilling their roles in one of the institutions that make up the basic structure of society, since racist incidents, in whatever context they occur, would clearly affect the quality of individuals’ relations with each other. Rather, it seems natural for it to hold in people’s everyday lives and therefore to conclude that it extends beyond the basic structure.

Finally, let us look briefly at the principle from which Rawls believes all obligations derive, which he calls “the principle of fairness”. This principle holds that

a person is required to do her part as defined by the rules of a legitimate institution if she has voluntarily made use of the institution to further her interests (TJ, p. 96).

Like the natural duty to support just institutions, this principle may seem to ask for nothing beyond the behaviour necessary to maintain the basic structure, and therefore not to demand anything in excess of what we would expect of a view that confined the scope of social justice to the basic structure. However, one principle that derives from the principle of fairness is not so restricted to the basic structure. What Rawls calls the “principle of fidelity” holds that promises that are made under the right conditions should be kept (TJ, pp. 303-306). This principle applies to all voluntary relationships that people enter into. It is part of a theory of social justice because of its role in enabling and stabilising co-operative arrangements for mutual advantage and in building trust among citizens. This latter consequence ties it to the development of the sense of justice, in ways discussed above for the natural duties.

In sum, the natural duties of respect and mutual aid, and the principle of fidelity cannot be reduced to part of the basic structure in the sense that they apply only to actions and attitudes of individuals while they are engaged in activities that take place within its institutions or that support its institutions. In addition, they are essential to establishing social justice as Rawls conceives it. It follows, contra Van Parijs (2003), that Rawls’s theory of social justice applies to more than just the basic structure. This argument gains further support from (and in turn supports) the interpretation of Rawls’s thought given in section 5.3. Indeed, the connection between these principles and the stability of a just society show that the second part of the quote from Rousseau applies to Rawls’s theory as well: if personal behaviour and motivation are excluded from the start, a well-ordered society will not be possible.

5.5 Further principles for individuals?

This conclusion lends more force to Cohen's initial question. For if principles for individuals are not excluded from Rawls's theory, why should it exclude a duty to act in ways that maximise the situation of the worst-off, subject to the constraints of some personal prerogative? We can also ask this question with respect to other goods than those regulated by the difference principle. There is at least an equally strong case for considering principles governing individuals' actions that influence the opportunities for education, skill acquisition, and jobs of the less well off, the kinds of attitudes people have towards each other in society,⁶⁴ and the functioning of the country's democracy.

Theda Skocpol (2004), for example, summarising recent research on the topic, argues that the kind of voluntary associations people join influences society's egalitarian culture and democratic participation. She documents a sharp decline over the last four decades in the United States in what she calls "membership organisations". These are groups like the labour unions and the General Federation of Women's Clubs, which emphasise and express solidarity among citizens who see themselves joined in shared moral undertakings, and which are largely run by voluntary workers and with a focus on active membership and face-to-face meetings. By contrast, there has been a marked increase in what she calls "professionally managed public interest associations". These are single-issue organisations like the Wilderness Society or the National Abortion and Reproductive Rights Action

⁶⁴ Such intangibles may fall under the difference principle, which applies also to the social bases of self-respect.

League, which are run almost exclusively by paid professionals, and which see members primarily as sources of financial contributions.

Membership organisations are joined by people from all class backgrounds. They cycle people from these backgrounds through official responsibilities, providing training in management and leadership skills along the way, and generally contribute to the interaction of members of all classes with a cross-section of the population. By contrast, professionally managed groups are joined and mostly run by members of highly educated classes, and have very low degrees of active membership participation. As a consequence, Skocpol believes, this shift has contributed to a decline in cross-class fellowship and lower-class democratic participation. If Skocpol's analysis is correct, and if membership organisations could not (or should not) be effectively promoted by the state, then there might be a case for requiring people to take these effects into account in choices about which kinds of voluntary associations they join and how the organisations they join should be run. As Skocpol writes: "In our activities as citizens, let each of us consider what we can do personally—by organising and joining together with our fellow citizens from all walks of life, and by imagining new modes of popular involvement that hold the promise of revitalising civic democracy" (2004, p. 15).

Rawlsian objections to further principles 1: publicity

Rawls proposes three reasons for not extending the range of principles for individuals or voluntary associations beyond those discussed in section 5.4. The first relates to considerations of publicity. Rawls's publicity constraint has two aspects (TJ),

pp. 48-49 and 114-115).⁶⁵ The first relates to its *content*. A rule satisfies the publicity constraint with regard to content when individuals know that this rule is the result of hypothetical contractual agreement; they know what it demands of them and of others,⁶⁶ and both these facts are common knowledge. The second relates to *compliance*. A rule satisfies the publicity constraint with regard to *compliance* when nearly everyone has reason to believe that nearly everyone complies with the rule, and nearly everyone has reason to believe that nearly everyone believes that nearly everyone complies. With regard to both aspects of the publicity condition, it is important to note that the ways in which we ascertain whether a rule satisfies the condition are restricted to methods of inquiry, sources of information and ways of reasoning that are generally accepted to be appropriate for questions of political or social justice (PL, pp. 66-67).

The reasons for which Rawls imposes this constraint are diverse.⁶⁷ The one that most concerns us here is related to the problem of assurance. Rawls assumes that individuals are motivated by an ideal of fair reciprocity: they are generally willing to keep to fair rules if they believe others are generally doing so, whilst their desire to stick to fair rules will be less strong or absent when others are not generally complying with it. It follows that publicity with regard to compliance, and hence with regard to content, is essential for individuals to comply with the demands of rules that might be considered as rules of justice, and therefore for the feasibility and stability of those rules.

⁶⁵ See also Vandebroucke (2001, pp. 225-227).

⁶⁶ Rawls sometimes describes this epistemic limitation as a part of what he calls the “universality constraint”. See TJ (p. 114).

⁶⁷ See PL (p. 66ff) and Vandebroucke (2001, pp. 262-276).

Now, at this point, Rawls appears to bring in a fact about human motivation, not under actual social conditions—since under actual conditions, people may not be generally motivated by fair reciprocity—but under ideal social conditions. (In this sense, what Rawls calls his “realistic utopianism”, or “probing of the limits of practical possibility” may best be captured not by the Rousseauian idea of “taking people as they are and laws as they might be” but as “taking people and laws as they might be”. See JF, p. 4.) It is a deep and difficult question whether in devising principles of justice, one should take facts of this kind as setting limits to which principles of justice one could adopt. Doing so would seem to rule out, for example, principles of justice which people could not generally be motivated to follow under ideal conditions. I will not comment on this question here, and will not question this aspect of the Rawlsian enterprise. I do, however, want to point out one consideration in support of the publicity condition that does not relate to facts about human psychology, but rather to the nature of what one is trying to achieve through one’s actions. For the aim of people who comply with the principles of justice is to achieve a just society, and this is something that cannot be achieved through one individual’s action alone. Moreover, if a sufficient share of the members of one’s society do not comply with the rules of justice designed for a society in which almost everyone complies with these rules, then complying with these rules oneself might not bring society any closer to justice. In such cases, reasonable assurance that there will be a sufficient degree of compliance with the requirements of justice is a prerequisite for being motivated to act on them, even if one is not motivated by reciprocity alone. It is noteworthy, however, that this does not apply to all principles of justice. If justice requires equality, for example, then promoting equality through one’s actions (say, by giving some of one’s money to a person who is less well off than the average person,

if one is better off than the average person) may bring society closer to justice, even if others do not adhere to just rules.

In what follows, I will simply take the publicity condition for granted. What does this condition mean for the role of principles for individuals? Williams (1998) has argued that only principles that apply to the basic structure meet the requirement of publicity. The discussion in sections 5.3 and 5.4 might seem to give us reason to question Williams's conclusion. For the principles for individuals that Rawls endorses go beyond the rules for the basic structure, and, as I will argue below, satisfy the demands of publicity. Though this might seem to contradict Williams's view, the conflict is, I believe, superficial. For Williams appears to employ a wider understanding of the "basic structure" than I have adopted, which includes not merely, as I have supposed, formal public rules, but also informal public rules.

To see that principles for individuals that do not apply to the basic structure (as I have understood the term throughout) alone can be public, let us take the content and compliance parts of publicity in turn. I think that the content condition must not be understood to require that individuals must on each occasion know exactly what it requires of them. For this may not be the case with rules that apply to the maintenance of just basic institutions, such as the requirement to vote for the party that one believes best upholds the principles of justice. There may be room for disagreement and individual judgement about which of the parties do so on a particular occasion, though the overall requirement is clear. Thus understood, this part of publicity demands simply that any rule should be able to be understood through paradigm cases and be able to provide general guidelines. Now, though the demands of the principles of mutual respect, mutual aid and the keeping of promises are not clear in every situation, they seem to be able to meet this requirement. The

principle of mutual aid, for example, demands that one help others who are in need when this is not too costly for oneself. Of course, it will not always be clear when someone is in need, or whether giving aid is so costly as to excuse us from aiding them. Nevertheless, there are clear cases where both kinds of conditions apply, and one can develop general guidelines on the basis of such cases. In addition, the reasoning required can be publicly shared, in the sense that it does not depend for its efficacy on keeping others in ignorance of one's principle of action, or on adopting any particular comprehensive moral doctrine.

The compliance condition can similarly be satisfied by norms that fall outside of the basic structure. For it is naturally understood as meaning that citizens have good grounds to believe that nearly everyone is generally complying with the rule based on the following kinds of information: firstly, the kinds of actions they see others perform and the attitudes they hear them express; secondly, the kinds of formal and informal sanctions they know apply in their society; and finally, their knowledge of human nature and the process of socialisation in their society (Vandenbroucke 2001, p. 276). It seems hardly more difficult to have this information and make judgements on the basis of it for individuals' general adherence to the rules pertaining to the institutions of the basic structure, such as paying taxes, than to do so for principles that apply to individuals' everyday lives such as those of mutual respect, aid and promise-keeping. Certainly, nothing about the nature of principles for individuals' everyday conduct rules out their being capable of meeting this condition.

Given that principles for individual conduct can meet the publicity condition, why does Rawls think that it rules out further rules for individual conduct? Rawls argues of certain principles at least—ones governing our economic choices with the aim of securing a just distribution of primary goods—that the information required

would be too great and complex to process. In *Political Liberalism*, Rawls addresses the question whether justice can be a product of individuals acting in accordance with some rules for individual conduct alone, without any need for specifically regulated basic institutions. To ensure that the results of individual actions are just, we need to know when the agreements individuals make with each other are freely made in the context of fair background conditions. Since the pattern arising from any initial position of freedom and fairness through unconstrained individual choice may lead to future situations of unfreedom and unfairness, if we consider justice as only the assignment of certain moral obligations to individuals, these obligations would have to take into account the social ramifications of individual choices. Rawls argues that this would be impracticable:

“There are no feasible rules that it is practicable to require economic agents to follow in their day-to-day transactions that can prevent these undesirable consequences. These consequences are often so far in the future, or so indirect, that the attempt to forestall them by restrictive rules that apply to individuals would be an excessive if not an impossible burden” (PL, p. 266).⁶⁸

The reason that there are no feasible rules is that such rules must not require too much information to be correctly applied. As Rawls writes, rules must therefore:

“not exceed the capacity of individuals to grasp and follow them with sufficient ease, nor (...) burden citizens with requirements of knowledge and foresight that they cannot normally meet. Individuals and associations cannot comprehend the ramifications of their particular actions viewed collectively, nor can they be expected to foresee future circumstances” (PL, p. 268).

⁶⁸ See also TJ (pp. 73-78, in particular 76).

Rawls believes that any principle enjoining people to take the effects on the situation of the worst off into account in their economic and personal choices must have these excessive requirements of knowledge, giving the example of the relatively simple case of bequests:

“It is obviously not sensible to impose on parents (as heads of families) the duty to adjust their own bequests to what they estimate the effects of the totality of actual bequests will be on the next generation, much less beyond” (PL, p. 267).

Rawls concludes that an “institutional division of labour” must be established between the basic structure and the rules applying directly to people’s choices in the economic sphere, which leaves individuals free to advance their ends within the framework of background institutions which carry out the operations required to maintain a just basic structure (PL, p. 284).

Now, this argument is directed at a view that only principles for individuals could create a just society. It is not clear how far its strictures apply to the case where we have a just basic structure supplemented by principles for individuals and associations. In the recent discussions of a duty to take the effects of one’s economic choices on the worst-off into account sparked by Cohen’s work, the necessary complexity of any such duty has been a major argument in defence of a Rawlsian division of labour between institutions and individuals and associations (see Williams, 1998 and Pogge, 2000). However, these discussions have not ruled out that, with sufficient ingenuity, and in particular social circumstances, a duty of this kind might meet the demands of publicity. Moreover, it appears to be relatively easy to think of other principles for individuals that would meet the demands of publicity. Devoting some of one’s time to participation in what Skocpol calls membership organisations

appears to be a case in point. In this case, the knowledge required to figure out what these duties demand would appear to be manageable. It would also seem possible for people to know the general degree of compliance with this duty. As another possible example of an additional principle that would not demand excessive information for individuals to be able to understand and follow, and that might meet the other demands of publicity, consider a requirement to take adequate care of one's health. If it is to everyone's advantage to have free universal health care coverage, as it saves on administrative costs and prevents intrusive information-gathering by public institutions, then individuals' choices that affect their health will influence the public cost of the health care system. We may think that this means that individuals should have a duty to take these external effects into account in their choices.⁶⁹ If they followed this duty, more resources would be available for other purposes. Again, the knowledge required to figure out what such a duty demands would appear to be manageable, as the current public campaigns on healthy living demonstrate. It would also seem possible for people to know the general degree of compliance with such a duty—news media already keep us well informed on the relevant trends.

Rawlsian objections to further principles 2: liberty

I now want to address a second objection to further principles for individual conduct that we might glean from Rawls's work: such principles would unacceptably limit people's liberty. We must note, however, that Rawls does not focus on how

⁶⁹ The current move in Britain towards public condemnation of unhealthy living, in part on the grounds that this places an unfair burden on the National Health Service, shows that many people think so.

principles for individual conduct limit liberty. Instead, he considers only coercive restrictions on individual behaviour and on voluntary associations. Rawls considers requiring democratic governance in churches with the aim of securing a more democratic culture (CP, p. 596), enforcing an equal division of household labour between men and women to secure fair opportunity (CP, p. 600), and imposing lump-sum taxes dependent on the assessment of people's earning capacity to get them to work at their most productive profession in order to improve the lot of the worst-off (CP, pp. 127 and 231; JF, pp. 64 and 157-158). In each case, Rawls argues that the restriction on the liberty of associations and individuals would be too great. The force of this argument appears diminished, however, by Rawls's focus on coercive intervention of the state as the way to achieve these aims. Clearly, coercive intervention by the state might undermine various basic liberties guaranteed by the liberty principle, such as freedom of association (which might be limited by the requirement of democratic governance in churches), and free choice of occupation (which might be limited by forcing people to work at their most productive profession).⁷⁰

⁷⁰ It is unclear whether Rawls conceives of freedom of occupation as being among the basic liberties. He never explicitly lists it as such, but some sections of *A Theory of Justice* appear to offer support for the thought that Rawls sees freedom of occupation as an important liberty (though perhaps not among the basic liberties). On TJ, p. 242, Rawls writes:

“The ideal scheme sketched in the next several sections makes considerable use of market arrangements. It is only in this way, I believe, that the problem of distribution can be handled as a case of pure procedural justice. Further, we also gain the advantages of efficiency and protect the important [not necessarily basic--AV] liberty of free choice of occupation.”

On TJ, p. 243 when defining the demands of formal equality of opportunity he writes:

However, as Cohen (2005) argues forcefully and in detail, an ethos that directs individuals who are unconstrained by government action to make choices that are to the benefit of the least well off does not undermine the freedom from being coerced to act in these ways. The liberty principle guarantees this freedom from coercion; an egalitarian ethos directs one to use this freedom in a particular way. Freedom of occupation, for example, is no more undermined by an ethos which requires that relatively advantaged individuals make their economic choices with the intention of maximising the prospects of the least advantaged, than one's political liberty is undermined by the duty, which Rawls endorses, to support the party that best realises the principles of justice (TJ, p. 294; see also Vandebroucke 2001, pp. 171-172).

Conclusion

None of the arguments for the restricted scope of the difference principle that we have reviewed are convincing. This leaves us with the original argument for an

“the government (...) also enforces and underwrites equality of opportunity in economic activities and in the free choice of occupation.”

And on p. 272:

“It is more important [than efficiency--AV] that a competitive scheme gives scope for the principle of free association and individual choice of occupation against a background of fair equality of opportunity (...).”

But it does appear as a powerfully expressed afterthought in *Justice as Fairness*, p. 64:

“The priority of liberty means that we cannot be forced to engage in work that is highly productive in terms of material goods. What kind of work people do, and how hard they do it, is up to them to decide in light of the various incentives society offers.”

egalitarian ethos: that its adoption will improve the situation of the worst-off and decrease inequality. In addition, our review of the role of the principles for individuals that Rawls does accept reveals another reason for favouring this ethos: its contribution to making social co-operation stable. This contribution runs through two routes. First, by diminishing inequality, it directly lessens the tendency for destructive envy to arise in society. Second, everyone's evident willingness to freely—that is, without coercion of the state—act to improve the prospects of the least advantaged is an affirmation of each individuals' worth, and especially of the worth of those individuals whose sense of self-worth might be undermined by being among the least advantaged in society. Thus, an egalitarian ethos would contribute to the stability of social co-operation through the same routes as the natural duties. An egalitarian ethos encompassing both the principles which Rawls identifies and further principles regulating individual conduct and motivation should therefore be part of an egalitarian theory of justice.

BIBLIOGRAPHY

Works by John Rawls

- CP *Collected Papers* edited by S. Freeman. Cambridge: Harvard University Press.
- JF *Justice as Fairness: A Restatement* edited by E. Kelly. Cambridge: Harvard University Press.
- PL *Political Liberalism* (2nd edition). New York: Columbia University Press.
- TJ *A Theory of Justice* (revised edition). Oxford: Oxford University Press.

Other references

- Anderson, E. (1999) 'What is the Point of Equality?', in *Ethics* 109, pp. 287-337.
- Arneson, R. (1989) 'Equality and Equal Opportunity for Welfare', in *Philosophical Studies* 56, pp. 77-93.
- Arneson, R. (1990a) 'Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare', in *Philosophy and Public Affairs* 19, pp. 158-194.
- Arneson, R. (1990b) 'Primary Goods Reconsidered' in *Noûs* 24, pp. 429-454.
- Arneson, R. (1990c) 'Neutrality and Utility' in *Canadian Journal of Philosophy* 20, pp. 215-240.
- Arneson, R. (1997) 'Postscript to Equality and Equal Opportunity for Welfare', in *Equality*. Edited by L. Pojman and R. Westmoreland. Oxford: Oxford University Press, pp. 238-241.
- Arneson, R. (1999) 'Egalitarianism and Responsibility', in *The Journal of Ethics* 3, pp. 225-247.

- Arneson, R. (2001) 'Luck and Equality', in *Proceedings of the Aristotelian Society*,
Supplementary Volume 75, pp. 73-90.
- Arneson, R. (forthcoming) 'Rawls, Responsibility, and Distributive Justice', in *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*. Edited by M. Salles and J. Weymark. Cambridge: Cambridge University Press. References are to the manuscript version of April 1997.
- Barry, B. (1989) *Theories of Justice*. London: Harvester-Wheatsheaf.
- Binmore, K. (1991) *Fun and Games*. Lexington, Mass: D.C. Heath.
- Binmore, K. and A. Voorhoeve (2003) 'Defending Transitivity Against Zeno's Paradox', in *Philosophy and Public Affairs* 31, pp. 272-279.
- Bossert, W. (1995) 'Redistribution Mechanisms Based on Individual Characteristics', in *Mathematical Social Sciences* 29, pp. 1-17.
- Bossert, W. and M. Fleurbaey (1996) 'Redistribution and Compensation', in *Social Choice and Welfare* 13, pp. 343-355.
- Bossert, W., M. Fleurbaey and D. Van de gaer (1999) 'Responsibility, Talent, and Compensation: A Second-Best Analysis', in *Review of Economic Design* 4, pp. 35-55.
- Brandt, R. (1979) *A Theory of the Good and the Right*. Oxford: Oxford University Press.
- Broome, J. (1991) *Weighing Goods*. Oxford: Basil Blackwell.
- Cohen, G. (1989) 'On the Currency of Egalitarian Justice', in *Ethics* 99, pp. 906-944.
- Cohen, G. (1993) 'Equality of What? On Welfare, Goods and Capabilities', in *The Quality of Life*. Edited by M. Nussbaum, and A. Sen. Oxford: Clarendon Press, pp. 9-29.
- Cohen, G. (1995) 'Incentives, Inequality, and Community', in *Equal Freedom. Selected Tanner Lectures on Human Values*. Edited by S. Darwall. Ann Arbor: University of Michigan Press, pp. 331-398.

- Cohen, G. (2000) *If You're an Egalitarian, How Come You're So Rich?* Cambridge: Harvard University Press.
- Cohen, G. (2003) *Rescuing Justice from Constructivism*. Mimeo, All Souls College, Oxford.
- Cohen, G. (2005) *The Freedom of, and in, Occupational Choice Trilemmas*. Mimeo, All Souls College, Oxford.
- Cohen, J. (2001) 'Taking People as They Are?', in *Philosophy & Public Affairs* 30, pp. 363-386.
- Daniels, N. (2003) 'Democratic Equality: Rawls's Complex Egalitarianism', in *The Cambridge Companion to Rawls*. Edited by S. Freeman. Cambridge: Cambridge University Press, pp. 241-276.
- Dworkin, R. (2000) *Sovereign Virtue*. Cambridge: Harvard University Press.
- Erasmus, D. (1960) *Brieven van Erasmus*. Translated and edited by O. Noordenbos. Utrecht: Prisma.
- Estlund, D. (1998) 'Liberalism, Equality and Fraternity in Cohen's Critique of Rawls', in *Journal of Political Philosophy* 6, pp. 99-112.
- Fleurbaey, M. (1994) 'On Fair Compensation', in *Theory and Decision* 36, pp. 277-307.
- Fleurbaey, M. (1995a) 'Equal Opportunity or Equal Social Outcome?', in *Economics and Philosophy* 11, pp. 25-55.
- Fleurbaey, M. (1995b) 'The Requisites of Equal Opportunity', in *Advances in Social Choice Theory and Cooperative Games*. Edited by W. A. Barnett, H. Moulin, M. Salles and W. Schofield. Cambridge: Cambridge University Press, pp. 37-53.
- Fleurbaey, M. (1995c) 'Equality and Responsibility', in *European Economic Review* 39, pp. 683-689.

- Fleurbaey, M. (1998) 'Equality Among Responsible Individuals', in *Freedom in Economics*. Edited by J. Laslier, M. Fleurbaey, N. Gravel, and A. Trannoy. London: Routledge, pp. 206-233.
- Fleurbaey, M. (2001) 'Egalitarian Opportunities', in *Law and Philosophy* 20, pp. 499-530.
- Frankfurt, H. (1982) 'The Importance of What We Care About', reprinted in his *The Importance of What We Care About*, Cambridge: Cambridge University Press, pp. 80-94.
- Frankfurt, H. (2004) *The Reasons of Love*. Princeton: Princeton University Press.
- Gourevitch, V. (1997) 'Introduction', in *Rousseau: The Social Contract and Other Later Political Writings*. Edited by V. Gourevitch. Cambridge: Cambridge University Press, pp. ix-xxxii.
- Hild, M. and Voorhoeve, A.. (2004) 'Equal Opportunity and Opportunity Dominance' in *Economics and Philosophy* 20, pp. 117-145.
- Hurley, S. (2004) *Justice, Luck, and Knowledge*. Cambridge: Harvard University Press.
- Knowles, E. (editor) (1999) *The Oxford Dictionary of Quotations*. Oxford: Oxford University Press.
- Kymlicka, W. (2002) *Contemporary Political Philosophy: An Introduction*. Oxford: Oxford University Press.
- Lewis, D. (1973) 'Causation', in *Journal of Philosophy* 70, pp. 556-567.
- Lippert-Rasmussen, K. (1999) 'Arneson on Equality of Opportunity for Welfare', in *Journal of Political Philosophy* 7, pp. 478-487.
- Lippert-Rasmussen, K. (2001) 'Egalitarianism, Option Luck, and Responsibility', in *Ethics* 111, pp. 548-579.

- Llavador, H. and J. Roemer (2001) 'An Equal-Opportunity Approach to the Allocation of International Aid', in *Journal of Development Economics*, 64, pp. 147-171.
- Meckled-Garcia, S. (2002) 'Why Work Harder? Equality, Social Duty, and the Market', in *Political Studies* 50, pp. 779-793.
- Munoz-Dardé, V. (2005) 'Equality and Division: Values in Principle', in *Proceedings of the Aristotelian Society, Supplementary Volume* 79, pp. 255-284.
- Norman, R. (1997) 'The Social Basis of Equality', in *Ratio* 10, pp. 238-252.
- Ooghe, E., E. Schokkaert, and D. Van de gaer (forthcoming) 'Equality of Opportunity Versus Equality of Opportunity Sets' forthcoming in *Social Choice and Welfare*. References are to the working paper from the Centre for Economic Studies, Katholieke Universiteit Leuven.
- Otsuka, M. (2003) *Libertarianism Without Inequality*. Oxford: Oxford University Press.
- Otsuka, M. (2004) 'Equality, Ambition, and Insurance', in *Proceedings of the Aristotelian Society, Supplementary Volume* 78, pp. 151-166.
- Phillips, A. (1999) *Which Equalities Matter?* London: Polity Press.
- Pogge, T. (2000) 'On the Site of Distributive Justice: Reflections on Cohen and Murphy', in *Philosophy & Public Affairs* 29, pp. 137-165.
- Rakowski, E. (1991) *Equal Justice*. Oxford: Clarendon Press.
- Raz, J. (1986) *The Morality of Freedom*. Oxford: Clarendon Press.
- Roemer, J. (1993) 'A Pragmatic Theory of Responsibility for the Egalitarian Planner', in *Philosophy & Public Affairs* 22, pp. 146-166.
- Roemer, J. (1996) *Theories of Distributive Justice*. Cambridge: Harvard University Press.
- Roemer, J. (1998) *Equality of Opportunity*. Cambridge: Harvard University Press.

- Roemer, J. (1999). *Equity in Health Care Delivery and Finance*. Mimeo, Department of Political Science, Yale University.
- Roemer, J. (2002) 'Equality of Opportunity: A Progress Report', in *Social Choice and Welfare* 19, pp. 455-471.
- Roemer, J. (2003) 'Defending Equality of Opportunity', in *The Monist* 86, pp. 261-282.
- Roemer, J. et al. (2003) 'To What Extent Do Fiscal Regimes Equalize Opportunities for Income Acquisition Among Citizens?', in *Journal of Public Economics* 87, pp. 539-565.
- Rousseau, J.J. (1988) *Political Writings*. Edited by A. Ritter and J. Conway Bondella. London: Norton.
- Scanlon, T. (1975) 'Preference and Urgency', reprinted in Scanlon *The Difficulty of Tolerance*. Cambridge: Cambridge University Press, pp. 70-83.
- Scanlon, T. (1991) 'The Moral Basis of Interpersonal Comparisons', in *Interpersonal Comparisons of Well-Being*, eds. Jon Elster and John Roemer. Cambridge: Cambridge University Press. pp. 17-44.
- Scanlon, T. (1995) 'The Significance of Choice', reprinted in *Equal Freedom: Selected Tanner Lectures on Human Values*. Edited by S. Darwall. Ann Arbor: University of Michigan Press, pp. 39-104.
- Scanlon, T. (1996) 'The Diversity of Objections to Inequality', reprinted in Scanlon *The Difficulty of Tolerance*. Cambridge: Cambridge University Press, pp. 202-218.
- Scanlon, T. (1998) *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Scheffler, S. (2003) 'What is Egalitarianism?', in *Philosophy & Public Affairs* 31, pp. 5-39.
- Sen, A. (2002) *Rationality and Freedom*. Cambridge: Harvard University Press.

- Skocpol, T. (2004) 'Voice and Inequality: The Transformation of American Civic Democracy', in *Perspectives on Politics* 2, pp. 3-20.
- Tawney, R. (1964) *Equality* (5th edition). London: Unwin.
- Temkin, L. (1993) *Inequality*. Oxford: Oxford University Press.
- Temkin, L. (2002) 'Equality, Priority, and the Levelling Down Objection'. In M. Clayton and A. Williams, editors, *The Ideal of Equality*. Palgrave, Basingstoke, pp. 126-161.
- Tressell, R. (1914) *The Ragged-Trousered Philanthropists*. Re-issued edition, 1997. London: Flamingo Publishers.
- Tungodden, B. (2003). 'The Value of Equality', in *Economics and Philosophy* 19: pp. 1-44.
- Vallentyne, P. (2002). 'Brute Luck, Option Luck, and Equality of Initial Opportunities', in *Ethics* 112, pp. 529-557.
- Van de gaer, D. (1993) *Equality of Opportunity and Investments in Human Capital*. Ph.D. Thesis, Katholieke Universiteit Leuven.
- Van de gaer, D., E. Schokkaert and M. Martinez (2001) 'Three Meanings of Intergenerational Mobility', in *Economica* 68, pp. 519-537.
- Van Parijs, P. (2003) 'Difference Principles', in *The Cambridge Companion to Rawls*. Edited by S. Freeman. Cambridge: Cambridge University Press, pp. 200-240.
- Vandenbroucke, F. (2001) *Social Justice and Individual Ethics in an Open Society*. Berlin: Springer Verlag.
- Voorhoeve, A. (2003) 'Harry Frankfurt on the Necessity of Love', in *Philosophical Writings* 23, pp. 55-70.
- Voorhoeve, A. and K. Binmore (forthcoming) 'Transitivity, the Sorites Paradox, and Similarity-Based Decision-Making', to appear in *Erkenntnis*.

Williams, A. (1998) 'Incentives, Inequality and Publicity', in *Philosophy & Public Affairs* 27, pp. 225-247.

Wolff, J. (2001) 'Levelling Down', in *Challenges to Democracy: The PSA Yearbook 2000*
Edited by K. Dowding, J. Hughes and H. Argetts. Palgrave, pp. 18-32.

Endnotes to Chapter 4.

ⁱ Let $1, \dots, k, \dots, K$ be a list of possible values of the non-responsible ethnicity variable and $1, \dots, l, \dots, L$ be a list of the possible values of the responsible education variable. In our example, $K = 3$ and $L = 5$. Unless mentioned otherwise, the indices k and l range over $\{1, \dots, K\}$ and $\{1, \dots, L\}$. I is a $K \times L$ matrix with incomes in the status quo depending on combinations of responsible and non-responsible characteristics and N is a $K \times L$ matrix with the number of the individuals possessing these combinations of characteristics. Policies are now of the form $\phi \in \mathbb{R}^{K \times L}$ and the set of feasible

policies satisfies the budget constraint $\sum_{kl} \max(\phi_{kl}, \rho \cdot \phi_{kl}) \cdot N_{kl} = 0$. Income under policy ϕ is

given by $I^\phi := I + \phi$. A negative transfer $r \in \mathbb{R}^-$ makes an amount of $\rho \cdot (-r)$ available for

redistribution. The contribution of an arbitrary transfer $r \in \mathbb{R}$ to the budget is therefore

$\min(-r, -\rho \cdot r) = -\max(r, \rho \cdot r)$. Since under our assumptions taxing individuals who are better off for a given level of education always increases the budget with which we can aid the less well off at that educational level, equal option sets need not always be achieved through levelling down.

Egalitarians maximise μ within the budget constraint $\sum_{kl} N_{kl} \cdot \max(\mu - I_{kl}, \rho(\mu - I_{kl})) = 0$.

Using non-linear optimisation algorithms, we find $\mu \approx 29,288$.

To determine the distribution which yields the most valuable option set, we note that given our assumptions, this is equivalent to maximising the average income of the population, consistent with equalising the return to education for different ethnicities for each level of education. This is equivalent to choosing the cheapest policy that is consistent with equalising the return to education for different ethnicities for each level of education. We can then show that the policy that does so is *self-financing in responsible characteristics*. A policy with the latter property equalises the income among individuals with the same responsible characteristic by using only transfers among these individuals. In other words, each group of individuals with the same level of education pays to redress the inequalities within its own ranks: college attendees pay for college attendees, and high school dropouts pay for high school dropouts. In proving this result, we make use of the concept of option-dominance

introduced in section 4.3. In search of the cheapest policy that is option-undominated (or OD-policy for short), our task is to minimise $B(\phi) := -\sum_{kl} \min(\phi_{kl}, 0) \cdot N_{kl}$ within the budget constraint

such that $(+)I_{kl}^\phi = I_{k'l}^\phi$ for all k, k', l . The budget constraint induces a non-linear boundary, thus complicating our optimisation problem. It is very helpful to note: *The unique solution to this optimisation problem is identical to self-financing in responsible characteristics.*

Proof: We note that $(++)$ neither $\phi \geq \phi'$ nor $\phi' \geq \phi$ for any $\phi, \phi' \in \Phi$ with $\phi \neq \phi'$.

Suppose that $\phi' \in \Phi$ is an OD-policy that minimises the budget $B(\cdot)$ and differs from the self-financing policy ϕ . Let $I_j^\phi := I_{1l}^\phi$ and $I_j^{\phi'} := I_{1l}^{\phi'}$ denote the income of individuals with responsible characteristic l under these two OD-policies. By $((++))$, there exists some l such

that $I_j^{\phi'} > I_j^\phi$ and some l' such that $I_{j'}^{\phi'} < I_{j'}^\phi$. Clearly, some of the l' -individuals must

receive negative transfers under ϕ' . We now define a policy ϕ'' that, first, agrees with ϕ' in

the treatment of all characteristics l'' that differ from both l and l' ($\phi''_{kl''} := \phi'_{kl''}$ for all

$l'' \neq l, l'$ and all k) and, second, agrees with ϕ in the treatment of characteristic l

($\phi''_{kl} := \phi_{kl}$ for all k). Compared to ϕ' , this policy redistributes

$$\delta := \sum_k \left[\max(\phi'_{kl}, \rho \cdot \phi'_{kl}) - \max(\phi_{kl}, \rho \cdot \phi_{kl}) \right] \cdot N_{kl}$$

less income among individuals

with characteristic l where $\delta > 0$. We finally use these savings to increase the equalised

income of individuals with characteristics l' (the values of $\phi''_{kl'}$ are fixed by $(+)$). Since some of

these payments go to l' -individuals who receive negative transfers under ϕ' , ϕ'' must have a

lower budget $B(\phi'')$ than ϕ' . Contradiction! ■

The condition that a policy be self-financing in responsible characteristics requires that

$$\sum_k \left[\max(\phi_{kl}, \rho \cdot \phi_{kl}) \right] \cdot N_{kl} = 0 \text{ for every } l. \text{ By } (+), \text{ we know that } \phi_{2l} = I_{1l} - I_{2l} + \phi_{1l}$$

and $\phi_{3l} = I_{1l} - I_{3l} + \phi_{1l}$. In the current example, a quick calculation shows that we must have

$\phi_{1l} < 0$ and $\phi_{2l}, \phi_{3l} > 0$. Self-financing then implies that

$$\phi_{1l} \cdot (N_{1l} \cdot \rho + N_{2l} + N_{3l}) + N_{2l}(I_{1l} - I_{2l}) + N_{3l}(I_{1l} - I_{3l}) = 0 \text{ where}$$

$$\phi_{1l} = \frac{N_{2l}(I_{2l} - I_{1l}) + N_{3l}(I_{3l} - I_{1l})}{N_{1l} \cdot \rho + N_{2l} + N_{3l}}.$$

ii Policies are now of the form $\phi \in \mathbb{R}^K$ and the set Φ of feasible policies satisfies

$$\sum_{1 \leq k \leq K} \max(\phi_k, \rho \cdot \phi_k) \cdot N_k = 0 \text{ where } N_k \text{ is the number of individuals with non-responsible}$$

characteristic k . (As an aside, this restriction of feasible policies means that all policies respect Bossert and Fleurbaey's (1996) axiom of 'equal transfer for equal non-responsible characteristics'.) Income

under policy ϕ for individuals with characteristics k and l equals $I_{kl}^\phi := I_{kl} + \phi_k$. In our example,

ϕ_1, ϕ_2, ϕ_3 represent transfers to Nationals, Immigrant Group 1, and Immigrant Group 2.

When computing the set of option-undominated policies by brute force, we have to search the entire set of feasible policies when testing whether a given policy is option-dominated. The running time required by the brute force algorithm increases quadratically in the size of the set of feasible policies. Although a more sophisticated algorithm has a somewhat better performance, the calculation still remains difficult. In the current model, it is, however, immensely simplified by the following useful equivalence. *A policy $\phi \in \Phi$ is undominated in the value of options if and only if (*) every ethnic group is worst-off at some level of education* (i.e. for all k there exists some l such that $I_{kl}^\phi = \min_{k'} I_{k'l}^\phi$).

Proof: (a) Suppose (*) is false. We have to show that ϕ is options-dominated. We know that there exists some k such that, for all l , $\delta_l := I_{kl}^\phi - \min_{k'} I_{k'l}^\phi \leq 0$. Let $\delta := \min_l \delta_l$. We can now redistribute $\rho \cdot \delta \cdot N_k > 0$ in a positive transfer among all individuals with non-responsible characteristics other than k , thus increasing their income. Hence, ϕ is option-dominated. (b) Suppose (*) is true and ϕ' differs from ϕ . We have to show that ϕ is option-undominated. By the budget constraint, there must exist some k such that $\phi'_k < \phi_k$. By (*),

there exists some l such that $I_{k'l}^{\phi} = \min_k I_{k'l}^{\phi}$. It follows that $I_{k'l}^{\phi'} < \min_k I_{k'l}^{\phi}$ and ϕ'

therefore does not option-dominate ϕ . ■

We note that this equivalence depends only on the following property of the set Φ of feasible policies: (**) Whenever $\phi, \phi' \in \Phi$ with $\phi \neq \phi'$, then neither policy dominates the other, in the sense that neither $\phi \geq \phi'$ nor $\phi \leq \phi'$. For feasible sets of this nature, it follows that the use of maximin instead of leximin is innocuous.

Redistributive cost is minimised exactly when $-\phi_1$ is minimised because all option-undominated policies (or OD-policies for short) have positive ϕ_2 and ϕ_3 (the latter is positive exactly if $-\phi_1 > \frac{N_2}{N_1 \cdot \rho} \cdot \phi_2$). By the equivalence proven above, an OD-policy ϕ renders Nationals at some educational level worst-off. Since the difference between Nationals' incomes and the smallest incomes is minimised at the lowest educational level, $-\phi_1$ is minimised when Nationals at this level obtain the same income as the rest of the population at that level. The incomes of individuals in this category are equalised precisely when $14885 + \phi_1 = 11948 + \phi_2 = 13129 + \phi_3$. Equivalently, we must have both $\phi_2 = 2937 + \phi_1$ and

$$\min\left(-\frac{(N_1 \cdot \rho + N_2) \cdot \phi_1 + 2937 \cdot N_2}{N_3}, -\frac{(N_1 \cdot \rho + N_2) \cdot \phi_1 + 2937 \cdot N_2}{N_3 \cdot \rho}\right) = 1756 + \phi_1.$$

This implies $\phi_1 \approx -501$, $\phi_2 \approx 2436$, $\phi_3 \approx 1255$. The resulting income shows that this policy is undominated in the value of options (by the above equivalence).

iii This rule has the unfortunate characteristic that one can increase the value of its target by making the worst off at a relatively low-valued combination of responsible characteristic even worse off, if by doing so the number of people who are worst off at that combination of responsible characteristics is decreased. For example, this rule evaluates the status quo, shown in table 1, at 22,635. Consider the alternative policy of reducing the income of members of Immigrant Group 2 at the lowest educational level by 1,182, making them worst off at this educational level. This money is not redistributed and no

further income transfers take place. Because there are fewer members of Immigrant Group 2 than there are members of Immigrant Group 1 at the lowest educational level, the weight of this level is now diminished. As a result, the above criterion evaluates this policy at 25,112, and thus prefers it to the status quo, even though it is option-dominated by the status quo. Cases like this are ruled out by the requirement of using Roemer's rule only to select among the set of option-undominated policies.

^{iv} When income transfers are based on ethnicity alone, we find

$\phi_1 \approx -1940$, $\phi_2 \approx 9907$ and $\phi_3 \approx 3660$, using algorithms for non-linear optimisation problems.

The resulting policy makes Nationals uniformly the worst off, and members of Immigrant Group 1 best off for all educational levels except the highest. Though this may at first seem surprising, the reason is that a 1 unit transfer from members of Immigrants from Group 1 to Nationals would make Immigrants from Group 1 uniformly the worst off at the education level 'BA or more' and thereby greatly decrease the weight of this education level in the overall calculation of the average welfare of the most disadvantaged. Since this education level yields the highest welfare, this will decrease the average welfare of the most disadvantaged.