

November 2009

Dear NYU Law, Economics, and Politics Colloquium participants:

What follows is a draft of chapter 4 from my forthcoming book, *Well-Being and Equity: A Framework for Policy Analysis* (Oxford U. Press 2010 (I hope!)). The first few pages of the chapter, including some long footnotes, provide a summary of material which is addressed in previous chapters and which is relevant to this chapter. These pages also give some sense of the overall aim of the book. Basically, the book tries to provide a systematic philosophical treatment and defense of the use of social welfare functions (SWFs) for policy analysis. For reasons covered in chapter 1, the book adopts the strategy of working within welfarism, rather than engaging familiar debates between welfarists and nonwelfarists. (The reason, in a nutshell, is that too little philosophical effort has been focused on figuring out the best specification of welfarism; and that the view is sufficiently plausible, particularly in the form of responsibility-adjusted welfarism, to merit fuller treatment). The construct of an SWF derives from theoretical welfare economics, and so the book ends up being an interdisciplinary work, trying to draw both from philosophy and from economics so as to defend and elaborate the SWF framework.

Chapter 1 covers preliminaries. It's a very quick review of metaethics and moral epistemology and an overview of the debates about welfarism. It also sets forth a formal, generic architecture for any welfarist choice-evaluation procedure; the SWF framework is, in turn, a particular version of that architecture. Chapter 2 argues that the SWF framework is a more attractive approach to welfarism than competing policy-analysis frameworks, which are currently dominant (cost-benefit analysis, various inequity metrics, cost-effectiveness analysis). Chapter 3 defends an account of well-being. Chapter 4, the chapter you have, argues for a prioritarian SWF. Chapter 5 addresses the "time slice" problem: should we be concerned about the distribution of lifetime or sublifetime well-being? Chapter 6 engages problems of estimation. The SWF requires interpersonally comparable utilities as its inputs. How shall we actually estimate those? Chapter 7 discusses the problem of applying an SWF under uncertainty and, in particular, the tension with ex ante Pareto indifference and ex ante superiority that arises with a prioritarian SWF. Chapter 8 addresses the problem of optimal legal institutions. Assuming that a prioritarian SWF is *morally* attractive, what should be its legal role?

I very much look forward to your comments. There is some substantive discussion in the footnotes, which may be worth looking at, but the citations are quite fragmentary and preliminary. This is a long chapter – apologies! It may make sense to skip over the section on utilitarianism, which is less central to the chapter than the subsequent sections – the sections that argue for a prioritarian SWF and against other kinds of distribution-sensitive SWFs. But I welcome a discussion of any part of the chapter or, for that matter, the broader book project. –
Matt Adler

Chapter 4: The Case for a Prioritarian SWF

This chapter argues that the most attractive social welfare function (“SWF”) has the form $w(u(x)) = \sum_{i=1}^N g(u_i(x))$, with $g(\cdot)$ strictly increasing and strictly concave. In particular, it has the

“Atkinsonian” form $w(u(x)) = \frac{1}{1-\gamma} \sum_{i=1}^N u_i(x)^{1-\gamma}$, with $\gamma > 0$. Given an outcome set \mathbf{O} and a set \mathbf{U}

of utility functions, a moral ranking of the outcome set should be generated using the rule: outcome x is morally at least as good as outcome y iff, for all $u(\cdot)$ belonging to \mathbf{U} ,

$$\frac{1}{1-\gamma} \sum_{i=1}^N u_i(x)^{1-\gamma} \geq \frac{1}{1-\gamma} \sum_{i=1}^N u_i(y)^{1-\gamma}. \quad \text{Or so I shall argue here.}$$

The reader will be reminded of where this chapter fits in the overall plan of this book. Chapter 1 situated my project within the terrain of contemporary metaethics, moral epistemology, and substantive moral philosophy. I work *within* welfarism, aiming to elaborate an attractive moral decision procedure which is welfarist in form. A moral decision procedure should be usable as an actual tool for morally evaluating governmental policies and other large-scale choices. A *welfarist* decision procedure has the following, generic, structure – one that is consequentialist and is focused on the well-being of persons. It derives a moral ranking (more precisely, a quasiordering) of a choice set \mathbf{A} from a moral ranking (more precisely, a quasiordering) of an outcome set \mathbf{O} , together with rules for choice under uncertainty. The outcome set, in turn, is associated with a set \mathbf{H} of life-histories: all pairings of an outcome and one of the N individuals in the population. An account of well-being produces a well-being quasiordering of the set of life-histories (as well as perhaps other well-being quasiorderings associated with the life-history set, for example a quasiordering of well-being differences between life-histories). And it is these well-being rankings of life-histories that help determine the moral ranking of the outcome set \mathbf{O} .¹

¹ See Chapter One for a full discussion of the generic structure of a welfarist decision procedure. To summarize very briefly: A quasiordering is a binary relation on a set of objects. It is reflexive and transitive; and is typically abbreviated using the symbol “ \succsim ”. The “at least as good as” relation is a quasiordering, and indeed throughout the book I often discuss various quasiorderings using the phrase “at least as good as.” “At least as good as” is reflexive: each item is at least as good as itself. It is transitive: If one item is at least as good as a second, and the second is at least as good as a third, then the first is at least as good as the third.

A complete quasiordering has the property that, for any pair of items, either the first is at least as good as the second, or the second is at least as good as the first. My analysis of the various quasiorderings involved in a welfarist decision procedure allows, but does not require, that they be complete. The device of a quasiordering is a rigorous and systematic way to allow for the possibility of *incomparability* in various rankings. If one item is not at least as good as the second, nor the second at least as good as the first, then the two are incomparable.

Given a quasiordering, we can define a “better than” relation, typically abbreviated “ \succ ”, as follows: one item is better than a second iff it is at least as good as the second item, and the second item is not at least as good as the first. Note that this relation is irreflexive, asymmetric, and transitive, which of course characterizes the concept of “better than.” We can also define an “equally good as” relation, abbreviated “ $=$ ”, as follows: one item is equally

good as a second iff it is at least as good as the second item, and the second item is at least as good as the first. *This* relation is reflexive, symmetric, and transitive, which characterizes the concept of “equally good as.”

A welfarist decision procedure provides moral guidance in choosing from a set **A** of possible choices. It does so by generating a moral quasiordering of the choice set. That quasiordering, in turn, derives in part from a moral quasiordering of an outcome set **O**. Outcomes are simplified possible worlds: cognitively tractable descriptions of possible realities. Under conditions of certainty, where each choice yields one outcome for certain, the connection between the choice and outcome set rankings is trivial: one action is at least as good as a second iff the outcome yielded by the first action is at least as good as the outcome yielded by the second. Under more realistic conditions of uncertainty, each action becomes a probability distribution over the outcome set; and the quasiordering of the outcome set is combined with this probability distribution to produce an ordering of the choice set. How to do so is controversial, and is the topic of Chapter 7.

The bulk of the book, prior to Chapter 7, focuses on how to construct a quasiordering of the outcome set. I assume that, in each choice situation, there is a fixed population of N individuals, identified by the numbers 1, 2, ... N , each of whom exists in each outcome. (What happens when this assumption of a fixed, finite population is relaxed? How should the SWF framework be employed in that case? That is a very important and very difficult question, highly relevant to problems of “future generations.” It is beyond the scope of this book to tackle this question. I have taken a stab elsewhere; see Adler (2009)).

A set of life-histories is constructed from the outcome set and fixed population of N individuals. Each life-history takes the form $(x; i)$, which means being individual i in outcome x . (Given a life-history $(x; i)$, I often refer to individual i as the “subject” of that life-history). Any welfarist decision procedure, as I see it, will have an account of well-being that produces a quasiordering of the life-history set. For any pair of life-histories, this account will yield the verdict (1) that the first life-history is at least as good as the second, and the second is at least as good as the first (i.e., that the two are equally good for well-being); or (2) that the first life-history is at least as good as the second, and the second is not at least as good as the first (i.e., that the first life-history is better than the second); or (3) that neither is true, in which case the life-histories are incomparable.

It bears emphasis that the quasiordering of the life-history set is a *well-being* quasiordering. If the account of well-being reaches the verdict that $(x; i)$ is at least as good as $(y; j)$, this means at least as good for the subject’s well-being – not “at least as good” in some other sense.

An account of well-being allows for *interpersonal* comparisons (more precisely, interpersonal comparisons of well-being levels) if there are some pairs of life-histories involving different subjects that are not incomparable. Some welfarist decision procedures associate further quasiorderings with the life-history set. In particular, the account of well-being developed in Chapter 3 constructs a *difference quasiordering* of the life-history set. This is a reflexive, transitive, binary relation between *pairs* of life-histories. Given one pair of life-histories $[(x; i), (y; j)]$ and another pair $[(z; k), (w; l)]$, the difference between $(x; i)$ and $(y; j)$ is greater than the difference between $(z; k)$ and $(w; l)$; or vice versa; or the differences are equal; or they are incomparable. An account of well-being allows for interpersonal comparisons of well-being differences if there is at least one case in which the difference between a pair of life-histories is comparable with the difference between a second pair of life-histories, and the subjects of all the life-histories are not identical.

A decision procedure is *welfarist*, as I see it, if the moral quasiordering of the outcome set satisfies certain basic “Pareto” constraints in terms of the quasiordering of the life-history set. One such constraint is Pareto indifference: If outcomes x and y are such that, for each individual i in the population, life-history $(x; i)$ and life-history $(y; i)$ are equally good, then outcomes x and y are equally morally good. The second such constraint is Pareto superiority, sometimes termed “strong Pareto”: If outcomes x and y are such that (1) for each individual i in the population, life-history $(x; i)$ is at least as good as life-history $(y; i)$, and in addition (2) there is at least one individual j such that life-history $(x; j)$ is better than life-history $(y; j)$, then outcome x is morally better than outcome y . Note that I am using the term “Pareto” to refer to relations between life-histories that have to do with *individual well-being*. If, in addition, well-being reduces to preference satisfaction, then the Pareto indifference condition as I have formulated it will be equivalent to the standard Pareto indifference condition in terms of preferences (i.e., each individual neither prefers outcome x to outcome y , nor vice versa); and similarly for the condition of Pareto superiority. However, I have framed the Pareto conditions in a more generic manner, so as to leave open what exactly the connection is between well-being and preferences.

As mentioned in the text below, it seems very plausible that a welfarist decision procedure – because it functions to provide moral advice --- should satisfy an impartiality constraint, in addition to Pareto indifference and superiority. If interpersonal level comparisons are possible, as I believe they are, impartiality can be elegantly

Any minimally plausible welfarist decision procedure should satisfy certain basic criteria. It should, indeed, contain a rule for generating a moral quasiordering of any outcome set. Further, at a minimum, this quasiordering should satisfy the very basic axioms of Pareto indifference and Pareto superiority.²

The SWF framework is a particular *kind* of welfarist decision procedure. A welfarist decision procedure follows the SWF format – as I conceptualize it – if the procedure has the following features. It allows not only for intrapersonal well-being comparisons of various kinds, but also for *interpersonal* comparisons of the well-being associated with life histories and/or interpersonal comparisons of the well-being differences between life histories. It represents all well-being comparisons via a set \mathbf{U} of utility functions, where each utility function $u(\cdot)$ in \mathbf{U} maps each life-history onto a utility number and each outcome onto an N -entry list or “vector” of utility numbers, one for each person in the population.³ Finally, the SWF framework ranks any given pair of outcomes, x and y , as a function of the set of utility vectors associated with each outcome via \mathbf{U} .⁴

Although SWFs are accepted by many theoretical welfare economists, and are employed in some bodies of economic scholarship to evaluate governmental policies – for example, in the field of optimal tax policy, and to some extent in environmental economics – other approaches to policy analysis are currently more widespread. In particular, cost-benefit analysis, cost-effectiveness analysis, and a variety of inequity metrics (inequality metrics, poverty metrics, social gradient metrics, and incidence metrics), are more widely used to evaluate governmental policies and other large-scale choices. Chapter 2 systematically surveyed these frameworks. The conclusion of Chapter 2, in short, was this: *if* interpersonal comparisons are possible, and *if* well-being comparisons can be represented via utility numbers, currently dominant non-SWF

captured in an “anonymity” constraint: if the pattern of well-being levels in outcome x is just a rearrangement (permutation) of the pattern of well-being levels in outcome y , then the two outcomes are equally morally good.

² See *supra* note 1.

³ In other words, each utility function in \mathbf{U} has two roles. To begin, and most fundamentally, it maps each life-history onto a single real number. In other words, it takes the form $u(x; i)$. The numbers assigned by each $u(\cdot)$ to the life-histories is then used to *represent* the well-being quasiordering of life-histories and/or the difference quasiordering. (For the particular representational rules that I employ, which are the most natural ones, see below note 6.)

However, each such scalar-valued $u(\cdot)$ can be immediately associated with a *vector-valued* utility function, which for simplicity I also abbreviate as $u(\cdot)$. The function $u(\cdot)$, in this sense, maps each outcome onto an N -entry list of utility numbers. That is: $u(x) = (u(x; 1), u(x; 2), \dots, u(x; N))$. In discussing $u(\cdot)$ in this sense, I will often abbreviate it as $u(x) = (u_1(x), u_2(x), \dots, u_N(x))$, where $u_i(x)$ equals $u(x; i)$. It is the utility assigned by $u(\cdot)$ to individual i in outcome x .

⁴ For a more precise formulation, see below, note 11. As discussed in Chapter 2, an SWF as I define it here – a rule that produces a quasiordering of outcomes as a function of a set of utility functions – is a generalization of the idea of a social welfare function employed in the existing literature. That literature typically sees an SWF as a rule that produces a complete ordering of outcomes as a function of a single utility function; and also investigates invariance requirements, demanding that the ordering be the same regardless of which utility function in some set is employed.

policy frameworks provide a less attractive basis for structuring a moral decision procedure than the SWF approach (or, alternatively, are simply variations on the SWF approach).⁵

Chapter 3 took up the challenge of articulating an account of well-being that allows for interpersonal comparisons and for the representation of well-being via utility numbers. It drew upon contemporary philosophical scholarship concerning the nature of well-being, as well as Harsanyi's work concerning "extended preferences. Given an outcome set, life-history set and a set of N individuals, each individual k can be associated with a set \mathbf{U}^k – the set of utility functions, derived using expected utility theory, that expectationally represent her *fully-informed, fully rational, self-interested extended preferences* regarding life-history lotteries, and that assign zero to a state of nonexistence. The set \mathbf{U} is simply the union of these individual sets. This account allow for inter- as well as interpersonal comparisons of well-being levels; for inter- as well as interpersonal comparisons of well-being differences; and for comparisons of life-histories to nonexistence. All such comparisons are represented by the utility functions in \mathbf{U} , via the natural rules summarized in the margin.⁶

The account of well-being I favor is *preferentialist*, but idealizes preferences (in ways described in Chapter 3). Further, pace Harsanyi, the account does not assume that individuals have the same ranking of life-histories and lotteries over life-histories. It allows for heterogeneity in individuals' (idealized, self-interested) extended preferences. Well-being, in turn, is understood as a matter of individuals' *convergent* (idealized, self-interested) extended preferences. Because the account appeals to convergent preferences, it can be characterized as both a preferentialist account of well-being, and an objective good account.

This chapter turns to the task of specifying an attractive SWF. It will draw both upon economic theory and on the philosophical literature. One possibility is a utilitarian SWF.

⁵ Actually, Chapter 2 reached a stronger conclusion. It concluded that some non-SWF frameworks are unattractive independent of the possibility of interpersonal well-being comparisons. But there are other frameworks that are plausible as "fallbacks," if it turns out that interpersonal well-being comparisons are not possible and/or not representable by utility numbers.

⁶ Life-history $(x; i)$ is at least as good as life-history $(y; j)$ iff, for all $u(\cdot)$ belonging to \mathbf{U} , $u(x; i) \geq u(y; j)$. The well-being difference between life-history $(x; i)$ and life-history $(y; j)$ is at least as great as the well-being difference between life-history $(z; k)$ and life-history $(w; l)$ iff, for all $u(\cdot)$ belonging to \mathbf{U} , $u(x; i) - u(y; j) \geq u(z; k) - u(w; l)$. Life history $(x; i)$ is at least as good as nonexistence iff, for all $u(\cdot)$ belonging to \mathbf{U} , $u(x; i) \geq 0$. The ratio between two life-histories $(x; i)$ and $(y; j)$ is between the values of r and s iff, for all $u(\cdot)$ belonging to \mathbf{U} , $r < u(x; i)/u(y; j) < s$.

These are not the only possible rules for representing well-being, but they are the most natural ones, and represent the dominant approach in the literature on measuring levels and differences. To see the possibility of a different rule, note for example that we might construct a set \mathbf{V} by taking each $u(\cdot)$ in \mathbf{U} and defining $v(\cdot)$ as $e^{u(\cdot)}$. Then $v(\cdot)$, via the following "less natural" rules, represents the very same well-being as $u(\cdot)$: Life-history $(x; i)$ is at least as good as life-history $(y; j)$ iff, for all $v(\cdot)$ belonging to \mathbf{V} , $v(x; i) \geq v(y; j)$. The well-being difference between life-history $(x; i)$ and life-history $(y; j)$ is at least as great as the well-being difference between life-history $(z; k)$ and life-history $(w; l)$ iff, for all $v(\cdot)$ belonging to \mathbf{V} , $v(x; i)/v(y; j) \geq v(z; k)/v(w; l)$. Life history $(x; i)$ is at least as good as nonexistence iff, for all $v(\cdot)$ belonging to \mathbf{V} , $v(x; i) \geq 1$. The ratio between two life-histories $(x; i)$ and $(y; j)$ is between the values of r and s iff, for all $v(\cdot)$ belonging to \mathbf{V} , $r < \ln v(x; i)/\ln v(y; j) < s$.

Another possibility is some kind of non-utilitarian SWF. Contemporary philosophical scholarship regarding the nature of equality is particularly helpful, here, in explicating a variety of plausible non-utilitarian moral views. One such view is “prioritarian”: an idea introduced into the literature by Derek Parfit, drawing in turn on scholarship by Thomas Nagel. A different non-utilitarian moral view, associated with the work of Larry Temkin, focuses on “comparative fairness.” Yet a different type of non-utilitarian moral view, defended by Roger Crisp, is “sufficientism.”

The chapter will begin by characterizing different types of SWFs. To begin, it will distinguish between utilitarian SWFs and SWFs that are sensitive to the distribution of well-being (an idea I will make precise). Within the family of distribution-sensitive SWFs, we can further distinguish between SWFs that satisfy a condition of separability and those that do not. Within the subfamily of distribution-sensitive, separable SWFs, some satisfy the Pigou-Dalton condition, while others do not. As I shall explain, “prioritarianism” – if that term is understood as a particular type of SWF -- corresponds to distribution-sensitive SWFs that are both separable and satisfy the Pigou-Dalton condition.

The chapter will then use this typology to structure my argument in favor of an SWF with the form $\sum_{i=1}^N g(u_i(x))$, $g(\cdot)$ strictly increasing and strictly concave. It is tedious to keep repeating this stipulation about the $g(\cdot)$ function; and so the reader should assume that, unless otherwise noted, whenever I use the formula “ $\sum_{i=1}^N g(u_i(x))$ ” or refer to the “ $g(\cdot)$ function,” that function is strictly increasing and strictly concave.

First, I will consider the choice between utilitarian SWFs and distribution-sensitive SWFs. Then, I will argue that *separable* distribution-sensitive SWFs are more attractive than non-separable distribution sensitive SWFs. Next, I will argue in favor of prioritarian SWFs, as against distribution-sensitive, separable SWFs that fail the Pigou-Dalton condition -- in particular a “sufficientist” SWF. The chapter will conclude by examining the different types of prioritarian SWFs. I will argue that the form $\sum_{i=1}^N g(u_i(x))$, which satisfies a continuity requirement, is more attractive than the leximin SWF⁷ and other prioritarian SWFs that fail the continuity requirement; and that the Atkinsonian form is in turn the preferred variant of the more

⁷ Some scholars use the term “prioritarian” to mean solely SWFs of the form $\sum_{i=1}^N g(u_i(x))$, by contrast with a leximin SWF. However, because the leximin SWF satisfies both the Pigou-Dalton condition and the separability condition, I classify it as one type of “prioritarian” SWF. See below.

general form $\sum_{i=1}^N g(u_i(x))$. For short, I will sometimes refer to prioritarian SWFs of the form

$\sum_{i=1}^N g(u_i(x))$ as “continuous prioritarian SWFs.”

Just as we can differentiate between various types of SWFs, so we can differentiate between various *justifications* that might be offered to argue in favor of a given SWF. One of the central themes in my discussion will be that there are two distinct conceptions of fairness that might justify a distribution-sensitive SWF: what I shall term an “across-outcome” conception of fairness, and what I shall term a “within outcome” conception, i.e., what Temkin terms “comparative fairness.” I shall argue that the “across-outcome” conception, suggested by Thomas Nagel’s work, immediately justifies a *separable* SWF and is best specified so as to justify a *prioritarian* (separable, Pigou-Dalton respecting) SWF. Further, I shall argue, the “across-outcome” conception flows naturally from welfarism. Welfarism, the across-outcome conception of fairness, and a prioritarian SWF fit together beautifully in reflective equilibrium⁸ – or so I shall contend.

Before I undertake the characterization of different types of SWFs and the argument in favor of a continuous, prioritarian SWF with the Atkinsonian form, a few preliminary notes are in order. To begin, the analysis and argumentation here is, to a substantial extent, *detachable* from the particular account of well-being offered in Chapter 3, and the particular methodology there employed for constructing a ranking of the differences between life-histories, representable by utilities – a methodology that employs expected utility theory to do this. All that *this* chapter presupposes is the following: there is an account of well-being that allows for inter- and intrapersonal comparisons of life-histories and differences between life-histories, as well as comparisons to a zero level; and such comparisons are represented via a set U , using the particular “natural” rules mentioned earlier.⁹ My defense of a prioritarian, Atkinsonian SWF does not hinge on the further claim that well-being is a matter of idealized, self-interested preferences; and it is fully consistent with a variety of methods for using any given account of well-being to generate an ordering of life-histories and differences and the utilities that represent these.¹⁰

⁸ Reflective equilibrium is the methodology for moral reasoning employed throughout this book. I argued in Chapter 1 that this methodology is robust with respect to a wide range of plausible metaethical positions.

⁹ Represented via the rules mentioned earlier: See note 6.

¹⁰ To be sure, the particular account of well-being defended in Chapter 3 figures in this book in many important ways. Constructing that account provides a concrete answer to skeptics about interpersonal comparisons, or skeptics about the possibility of measuring well-being differences. Further, how to *implement* the SWF framework depends centrally on the particular account of well-being employed. Chapter 6 discusses how to use survey and other data to estimate utility functions; and my analysis of well-being in terms of idealized, self-interested extended preferences is the foundation for Chapter 6.

Throughout the chapter and indeed the book, I refer sometimes to “the SWF framework” and sometimes to “an SWF.” By the former, I mean the whole package of an account of well-being that allows for inter- as well as intrapersonal comparisons between life histories; a set \mathbf{U} of utility functions that represents the well-being associated with life-histories; and a rule that produces a quasiordering of an outcome set, depending on the sets of utility vectors associated with each outcome. By an “SWF,” I mean this rule. (A more precise characterization is given in the margin.)¹¹ Sometimes, the rule will incorporate a real-valued mathematical function. For example, the rule might be: x is at least as good as y iff $w(u(x)) \geq w(u(y))$ for all $u(\cdot)$ belonging to \mathbf{U} . However, the construct of an SWF, as I use it here, is more general. The leximin SWF, in particular, incorporates a rule for ranking utility vectors which does not employ a real-valued mathematical function.

I assume that any minimally plausible SWF will satisfy the minimal welfarist criteria outlined in Chapter 1: it will yield a quasiordering of any given outcome set, one that respects the basic welfarist principles of Pareto indifference and Pareto superiority. In addition, because the SWF framework incorporates an account of well-being that allows for interpersonal comparisons of well-being levels, we can formulate a requirement of *anonymity* that any plausible SWF will satisfy: if the distribution of well-being in outcome x is identical to the distribution in outcome y , save for the names of the individuals involved, then the outcomes are equally good.

But the case for a continuous prioritarian SWF presented in this chapter, as well as the analyses of the time-slice problem in Chapter 5, and of the application of an SWF under uncertainty in Chapter 7, are independent of the particular account of well-being defended in Chapter 3.

¹¹ Given any pair of outcomes x and y , each $u(\cdot)$ belonging to \mathbf{U} associates x with an N -entry utility vector and y with an N -entry utility vector. An SWF, as I mean it, ranks x and y (such that x is at least as good as y , or y at least as good as x , or both, or neither), depending just on the set of pairs of vectors assigned to x and y by all the utility functions in \mathbf{U} . (This is a kind of “independence of irrelevant alternatives” condition, and rules out more complicated possibilities: that the ranking of x and y depends on the utilities assigned to other outcomes; or that it depends on information about outcomes other than the utilities assigned them by the functions in \mathbf{U}).

An SWF, thus defined, need not have a “supervaluationist” form. But the simplest way to construct an SWF that produces a quasiordering is via the “supervaluationist” form, and all the SWFs considered in this chapter in fact have that form. In other words, they contain some rule R for ranking pairs of utility vectors; and they then say: x is at least as good as y iff $u(x)$ is at least as good as $u(y)$ according to R , for all $u(\cdot)$ belonging to \mathbf{U} .

In addition, as a matter of consistency across difference choice situations with the same population size N , it is very plausible to think that R should take the form of a single rule for ranking all possible N -entry utility vectors (or at least all such vectors with nonnegative utilities), which is then applied in any outcome set with N individuals – regardless of the particular outcomes in the set or the particular set \mathbf{U} . This chapter focuses on SWFs of this sort. This is very much in keeping with existing theoretical scholarship on social welfare functions, which also focuses on different rules for ranking all possible N -dimensional utility vectors.

It would be nice to have SWFs that are not merely consistent across difference choice situations with the same population size, but also consistent across choice situations with different sized populations. How to precisely formulate such a consistency requirement is not clear, and I will not attempt to do so. Intuitively, using the rule I

recommend -- x is at least as good as y iff $\sum_{i=1}^N g(u_i(x)) \geq \sum_{i=1}^N g(u_i(y))$ for all $u(\cdot)$ belonging to \mathbf{U} – in every choice

situation, with the very same $g(\cdot)$ function, regardless of the size N of the population, the outcome set, and the particular elements of \mathbf{U} , is fully consistent across choice situations.

Let us say that an SWF which satisfies these requirements is “Paretian” and “anonymous.” What this chapter does is to clarify the *further* axioms that a Paretian, anonymous SWF might be asked to satisfy -- namely, separability, the Pigou-Dalton principle, and a continuity requirement – and to engage in substantive moral argumentation to sort between these axioms.

Characterizing SWFs

How shall we usefully divide up the universe of Paretian anonymous SWFs? In this section, I shall be focusing on describing different types of SWFs – different types of rules for ranking the utility vectors associated with outcomes – rather than on describing different justifications for an SWF.

The rationale for any typology of SWFs is ultimately epistemic and pragmatic. Such a typology should be helpful in the “reflective equilibrium” process of identifying the most attractive SWF to be employed in a welfarist decision procedure that has the SWF format. This epistemic criterion is, in turn, quite fuzzy, and so dogmatism about one or another typology being the best seems out of place. Still, the following categorization scheme strikes me as particularly useful.

An initial distinction is between the utilitarian SWF and SWFs that are sensitive to the distribution of well-being. The utilitarian SWF sums unweighted utilities. It has the following form:

The Utilitarian SWF: Outcome x is at least as good as outcome y iff, for all $u(\cdot)$

$$\text{belonging to } \mathbf{U}, \sum_{i=1}^N u_i(x) \geq \sum_{i=1}^N u_i(y).^{12}$$

By “distribution sensitive” SWFs, I mean to characterize a very large family of non-utilitarian SWFs: all the SWFs that are plausibly justified in light of *some* non-utilitarian welfarist view that has substantial philosophical support. I use the neutral term “distribution sensitive” rather than “egalitarian,” because – at least these days – “egalitarianism” is often seen as a competitor to prioritarianism. Since some non-utilitarian SWFs are wildly implausible – for example, “leximax,” the “evil twin” of leximin, which gives lexical priority to better-off persons¹³ – it is helpful to have some criterion for distinguishing between the non-utilitarian SWFs that are worth debating, and others. That is what the idea of “distribution sensitivity” seeks to accomplish.

Economists tend to assume that any plausible non-utilitarian view will accept the Pigou-Dalton principle, but that is clearly wrong as a substantive matter. (As we’ll see, Temkin’s notion of comparative fairness might well justify a non-utilitarian SWF that rejects the Pigou-

¹² This is just the natural, “supervaluationist,” generalization, to an entire set \mathbf{U} , of the standard utilitarian SWF defined as the unweighted sum of individual utilities.

¹³ See Bossert & Weymark (2004; 1113).

Dalton principle.) A less demanding requirement is a preference for perfect equality. Even that is a bit too strong,¹⁴ but I think the following characterization does work well to identify the subset of non-utilitarian SWFs that are “distribution sensitive”. A *distribution-sensitive SWF* is such that: if overall well-being is the same in x and y , and the distribution of well-being in x is perfectly equal, but not in y , then x is at least as good as y .¹⁵

Within the large family of distribution-sensitive SWFs, we might distinguish, first, between distribution-sensitive SWFs that satisfy a separability-across-persons requirement and those that do not. Separability-across-persons means that the well-being of unaffected individuals¹⁶ does not affect the ranking of outcomes.

Separability-across-persons: If there is a group of individuals each of whom is equally well off in outcome x as she is in outcome y , then the moral ranking of x and y does not depend upon what those individuals’ well-being levels are.¹⁷

It bears noting that the idea of the separability of the ranking of outcomes, life-histories, or choices with respect to some “dimension” of their goodness is a very general idea. We can, at a minimum, talk meaningfully about, and substantively debate, the separability of the moral ranking of outcomes with respect to times, persons, or attributes; the separability of the well-being ranking of life-histories with respect to times or attributes; and the separability of the moral ranking of choices with respect to times, persons, attributes, or states of nature. All these sorts of separability have been discussed in economics and philosophy. In this chapter, as a shorthand, I often refer to the separability-across-persons axiom as, simply, “separability.” But it must be remembered that the focus of this chapter is separability in the moral ranking of

¹⁴ In particular, the sufficientist SWF, discussed below, fails to prefer perfect equality. If all individuals in outcome x are above the threshold, and well-being is unequally distributed, then outcome y which has the same total well-being as x and a perfectly equal distribution is counted by the sufficient SWF as equally morally good as x , not better.

¹⁵ Overall well-being is the same in x and y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $\sum_{i=1}^N u_i(x) = \sum_{i=1}^N u_i(y)$. Well-being is

perfectly equally distributed in x iff, for all $u(\cdot)$ belonging to \mathbf{U} , and for all pairs of persons, i and j , $u(x; i) = u(x; j)$.

¹⁶ Throughout the chapter, I will say that an individual is “unaffected” with respect to two outcomes if she is equally well off in the two, and “affected” if she is not equally well off in both (either better off in one or incomparably well off in one).

¹⁷ Because the SWF framework uses a set \mathbf{U} of utility functions to represent well-being, separability-across-persons can be given an equivalent formulation in terms of utilities. Consider two outcomes x and y . Assume there is an “unaffected” subgroup of the population of N individuals, such that for each individual k in this subgroup, $u(x; k) = u(y; k)$ for all $u(\cdot)$ belonging to \mathbf{U} . (Note that each k is equally well off in the two outcomes iff this is the case.) Imagine now that each $u(\cdot)$ belonging to \mathbf{U} is replaced by some function $u^*(\cdot)$, which has the property that $u^*(x; i) = u(x; i)$ and $u^*(y; i) = u(y; i)$ for each individual i who is not in the subgroup. So the set \mathbf{U}^* of these new functions contains exactly the same utility information regarding the well-being of affected individuals in the two outcomes as the original \mathbf{U} . Further, let us stipulate that the set \mathbf{U}^* of these new utility functions is such that individuals in the unaffected subgroup remain unaffected. In other words, for each individual k in the subgroup, $u^*(x; k) = u^*(y; k)$ for all $u^*(\cdot)$ belonging to \mathbf{U}^* .

If some \mathbf{U}^* is constructed from \mathbf{U} in this manner, separability-across-persons requires that the SWF rank x at least as good as y with \mathbf{U} in hand iff it ranks x at least as good as y with \mathbf{U}^* in hand.

outcomes, not the ranking of life-histories or actions; and that what I mean to defend here is the separability of the moral ranking of outcomes with respect to *persons*, not times or attributes.

Second, we might distinguish between distribution-sensitive SWFs that satisfy the Pigou-Dalton principle, and those that do not. The Pigou-Dalton principle was discussed in Chapter 2, in connection with inequality metrics. As noted there, the principle can be formulated in terms of different “currencies”: in terms of income, health, happiness, well-being, etc. I am interested, in this chapter, in the Pigou-Dalton principle in terms of well-being (which I’ll henceforth refer to simply as the Pigou-Dalton principle). What it says is that a pure transfer of well-being, from a better-off to a worse-off individual, which leaves everyone else unaffected, and leaves the worse-off individual still worse off,¹⁸ yields a better outcome.

Pigou-Dalton principle in terms of well-being: If individual i is worse off than individual j in both outcomes x and y ; individual i is better off in y than x ; individual j is better off in x than y ; the difference between individual i ’s well-being in y and x is the same as the difference between individual j ’s well-being in x and y ; and everyone else is just as well off in outcome x as in outcome y ; *then* y is a better outcome than x .¹⁹

It should be stressed that the Pigou-Dalton principle and separability are logically distinct. It is possible to formulate distribution-sensitive SWFs that satisfy the Pigou-Dalton principle but not separability. For example, a “rank-weighted” SWF, which corresponds to the Gini coefficient – a widely used measure of inequality – is a non-separable SWF that satisfies the Pigou-Dalton principle.²⁰

¹⁸Occasionally this principle is formulated in a broader way, requiring that a pure transfer which makes the worse-off individual better off, but shrinks the gap between his well-being and the other individual, be an improvement. It is not hard to show that any rule R for ordering all N -dimensional utility vectors (or all nonnegative such vectors) must satisfy this broader principle if it satisfies the narrower one plus an anonymity requirement; and thus that all SWFs of the form considered in this chapter will satisfy the broader principle if they do the narrower.

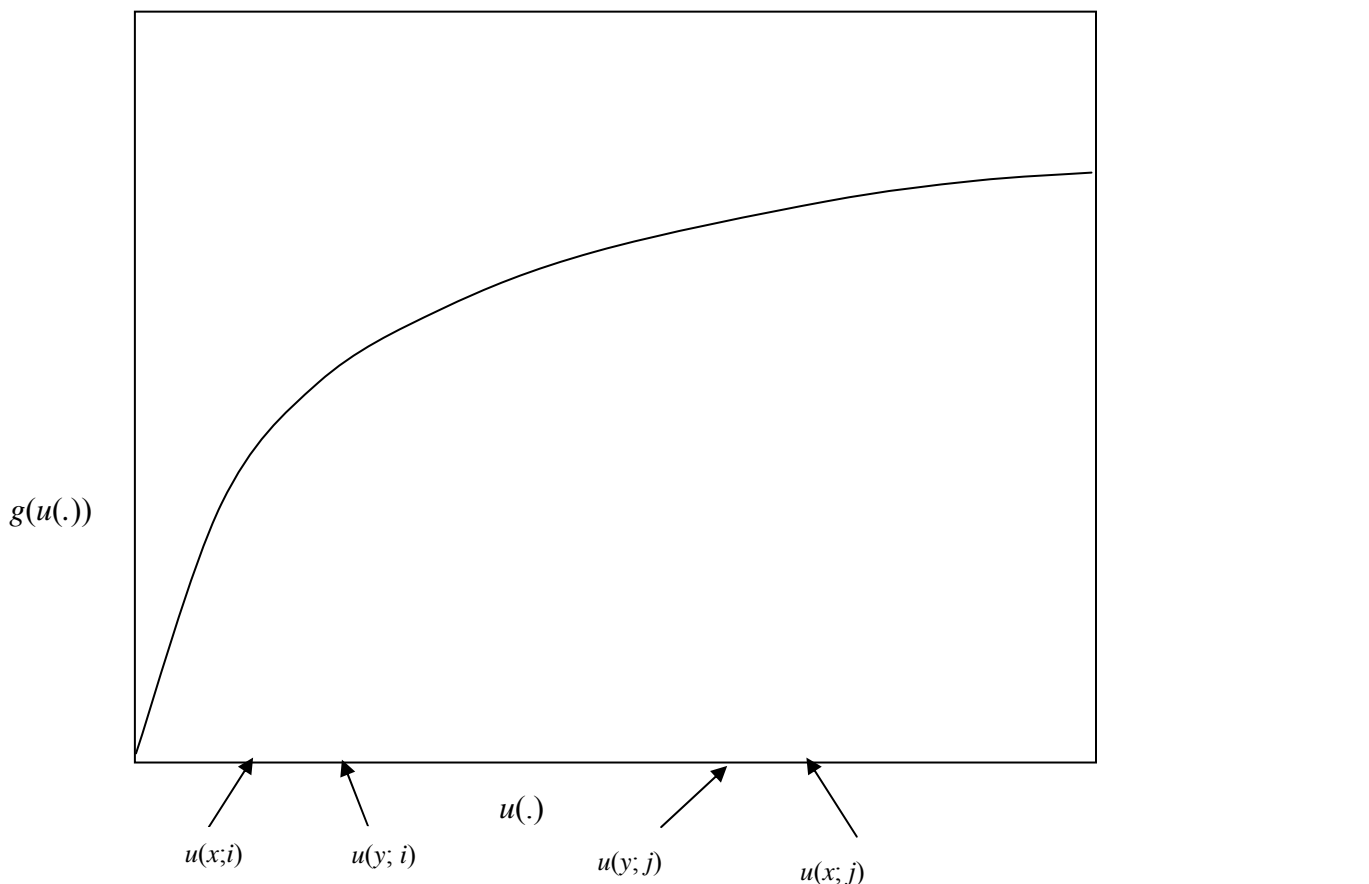
¹⁹ Once more, with a set \mathbf{U} in hand, this can be given an equivalent formulation in terms of utilities. Consider two outcomes x and y , such that everyone except i and j is just as well off in both outcomes. In other words, for each individual k , k distinct from i and j , $u(x; k) = u(y; k)$ for all $u(\cdot)$ belonging to \mathbf{U} . Imagine further that, for all $u(\cdot)$ belonging to \mathbf{U} , the following hold true: (1) $u(x; i) \leq u(x; j)$ and $u(y; i) \leq u(y; j)$, with each inequality strict for at least one $u(\cdot)$; (2) $u(y; i) - u(x; i) = u(x; j) - u(y; j) > 0$; and (3) $u(y; i) \geq u(x; i)$ and $u(x; j) \geq u(y; j)$, with each inequality strict for at least one $u(\cdot)$. The first condition is equivalent to saying that i is worse off than j in both outcomes; the second, to saying that the (nonzero) well-being differences between the two outcomes are the same for both individuals; the third, to saying that individual i is better off in y and individual j in x . If all this is true, then the Pigou-Dalton principle requires that y be ranked by the SWF as morally better than x .

²⁰ A rank-weighted rule assigns the lowest ranked utility a fixed weight, the next-lowest utility a smaller fixed weight, and so forth, and then sums the thus-weighted utilities. Formally, there are N weights, a_1 through a_N , which are strictly decreasing. For a given utility vector, reorder its utilities from lowest to highest, splitting ties arbitrarily. Assign the vector a score equaling a_1 times the lowest utility, plus a_2 times the second-lowest, plus ... plus a_N times the highest. The simplest rank-weighted rule uses weights $N, N-1, \dots, 1$.

A rank-weighted SWF employs some rank-weighted rule and says: x is at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $u(x)$ is at least as good as $u(y)$ according to that rank weighted rule. This SWF will not satisfy separability. (This is because changing the ranks of unaffected individuals can change the ranks of affected individuals, and thus the weighting of their well-being changes. See below for an illustration.) However, the rank-

Distribution-sensitive SWFs that satisfy separability but not Pigou-Dalton, or distribution-sensitive SWFs that fail both separability and Pigou-Dalton, are also possible.²¹ Such SWFs have not played much of a role in economic scholarship, but they clearly are logically possible.

Finally, of course, it is possible to have distribution-sensitive SWFs that satisfy both separability and Pigou-Dalton. My favored SWF does so. This SWF, again, says: outcome x is at least as good as outcome y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $\sum_{i=1}^N g(u_i(x)) \geq \sum_{i=1}^N g(u_i(y))$.²² In other words, it has an additive form, ranking a given pair of utility vectors by summing individual utilities transformed by the $g(\cdot)$ function. This SWF satisfies separability. Moreover, because the $g(\cdot)$ function is strictly concave, this SWF satisfies the Pigou-Dalton principle.



weighted SWF does satisfy the Pigou-Dalton principle. On this, and the connection between rank-weighted SWFs and the Gini coefficient, see Blackorby et al. (2005; 99-100); Adler & Sanchirico (2006; 301-02, 368-69).

²¹ The sufficientist SWF, described later in the chapter, is a distribution sensitive SWF which is separable but fails Pigou-Dalton. The failure of Pigou-Dalton occurs because it employs a threshold, above which pure transfers of well-being have no moral impact. It is easy to construct a rank-weighted SWF with a threshold that fails both separability and Pigou Dalton (but would still be distribution sensitive). Analogously to the “sufficientist” approach described below, construct a two-stage approach that compares two utility vectors by transforming each into a below- and above-threshold vector; applying the rank-weighted rule to the below-threshold vectors; and if the two vectors are ranked equally good at this stage, comparing unweighted sums of the above-threshold utility vectors.

²² The Atkinsonian form is a further specification of this.

Note that the difference between $u(y; i)$ and $u(x; i)$ is the same as the difference between $u(x; j)$ and $u(y; j)$, but, because $g(\cdot)$ is concave, the difference between $g(u(y; i))$ and $g(u(x; i))$ is greater than the difference between $g(u(x; j))$ and $g(u(y; j))$, meaning that the Pigou-Dalton principle is satisfied if vectors are compared by summing utilities transformed by the $g(\cdot)$ function

Finally, because the formula $\sum_{i=1}^N g(u_i(x))$ takes account solely of individual utility numbers,

ignoring proper names or any other information, it satisfies the very basic requirement of Pareto indifference; and because $g(\cdot)$ is strictly increasing, it satisfies the very basic requirement of Pareto superiority.

However, my favored SWF is *not* the only SWF that satisfies both separability and Pigou-Dalton. A different approach is the leximin SWF, which uses the leximin rule for ordering utility vectors, widely discussed in economic theory.

The Leximin SWF

The leximin rule for ordering utility vectors, $u(x)$ and $u(y)$, says this. If two vectors are permutations of each other, they are equally good. Otherwise, take each vector and reorder its utilities from smallest to largest, splitting ties arbitrarily. Call the reordered vectors $u^*(x)$ and $u^*(y)$. Find the lowest value of m , such that the m th entry of $u^*(x)$ is not equal to the m th entry of $u^*(y)$. If the m th entry of $u^*(x)$ is greater than the m th entry of $u^*(y)$, $u(x)$ is ranked higher by the leximin rule than $u(y)$; otherwise $u(y)$ is ranked higher.

The leximin SWF says: outcome x is at least as good as outcome y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $u(x)$ is at least as good as $u(y)$ according to the leximin rule.²³

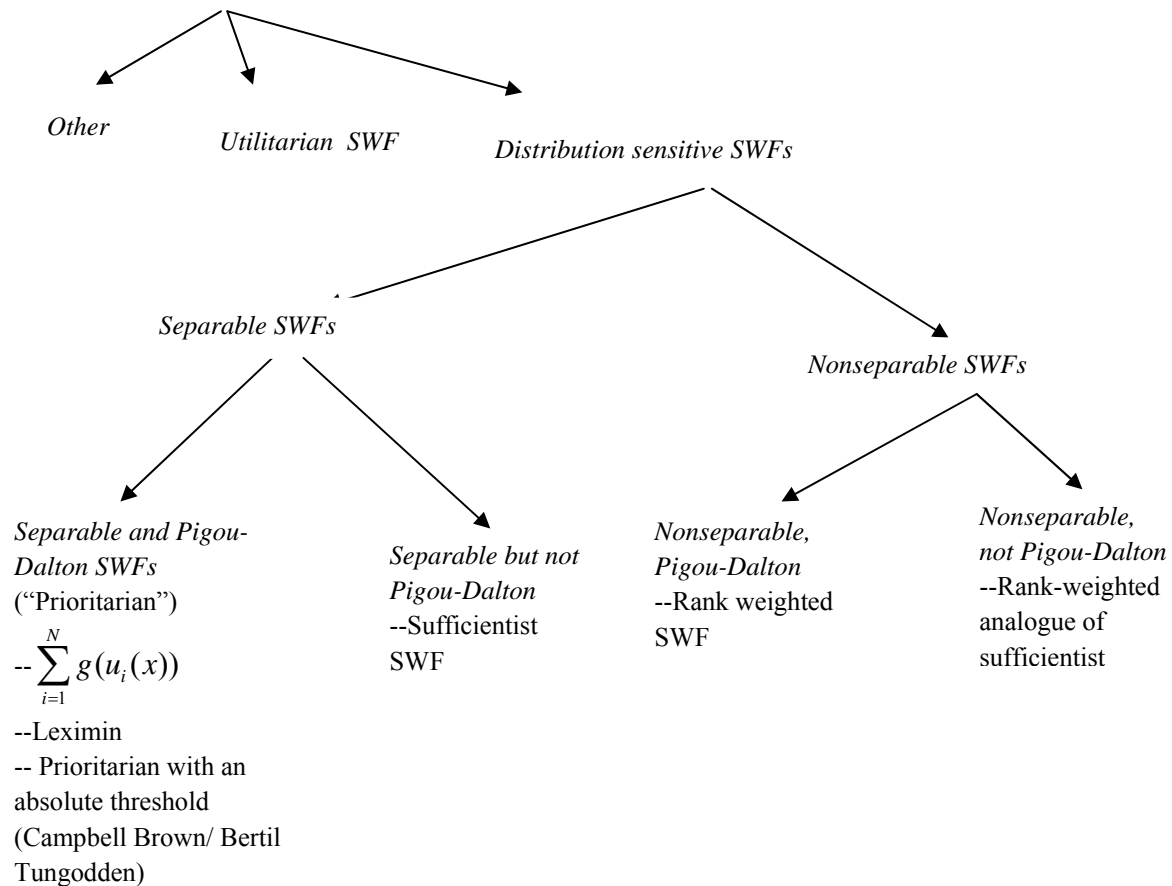
An SWF proposed by the philosopher Campbell Brown, as a way to remedy certain deficits in “sufficientism,” also satisfies the Pigou-Dalton principle and separability.²⁴ My favored SWF satisfies a continuity requirement, while these competitors do not.

The following diagram illustrates the typology of SWFs that we have arrived at.

²³ “At least as good” meaning either equally good according to the leximin rule, or better according to the leximin rule.

²⁴ See below.

A Typology of SWFs



How, now, does this typology relate to prioritarianism?

The recent, intense philosophical interest in “prioritarianism” was triggered by Derek Parfit’s 1991 Lindley Lecture, “Equality or Priority.”²⁵ In that lecture, Parfit discusses at length a hypothetical case described by Thomas Nagel in Nagel’s 1977 Tanner Lecture on equality.²⁶ The case, as Nagel describes it, is as follows:

Suppose I have two children, one of which is normal and quite happy, and the other of which suffers from a painful handicap. Call them respectively the first child and the second child. I am about to change jobs. Suppose I must decide between moving to an expensive city where the second child can receive special medical treatment and schooling, but where the family’s standard of living will be lower and the neighborhood will be unpleasant and dangerous for the first child -- or else moving to a pleasant semi-rural suburb where the first child, who has a special interest in sports and nature, can have a free and agreeable

²⁵Parfit (1991).

²⁶ Nagel (1977).

life. ... I want to suppose that the case has the following feature: the gain to the first child of moving to the suburb is substantially greater than the gain to the second child of moving to the city. After all, the second child will also suffer from the family's reduced standard of living and the disagreeable environment. And the educational and therapeutic benefits will not make him happy but only less miserable. For the first child, on the other hand, the choice is between a happy life and a disagreeable one.²⁷

Parfit, discussing this example,²⁸ supposes it can be represented using interpersonally comparable numbers measuring the well-being of the two children, as follows:

	<u>Move to city</u>	<u>Move to suburb</u>
First child	20	25
Second child	10	9

I will change the example slightly, by using a formula such as “ $25c$ ” to represent each child's well-being level, meaning that \mathbf{U} consists of a single utility function and all positive multiples thereof (in other words, that utility is unique “up to a positive ratio transformation”). This itself is a simplification, but a more realistic one than using a single utility number to represent someone's well-being in an outcome, which my account of well-being sees as implausible.²⁹

	<u>Move to City</u>	<u>Move to Suburb</u>
First child	$20c$	$25c$

²⁷*Id.* at 123-24.

²⁸ Parfit (1991; 81-83).

²⁹To be precise: let us say that a utility function $v(\cdot)$ is a positive multiple of utility function $u(\cdot)$ if there exists a positive number c , such that $v(x; i)$ is equal to $cu(x; i)$ for all life histories $(x; i)$. Then utility is unique up to a positive ratio transformation iff there is a utility function $u(\cdot)$ such that every positive multiple of $u(\cdot)$ is in \mathbf{U} , and a utility function is in \mathbf{U} only if it is a positive multiple of $u(\cdot)$.

The account of well-being defended in Chapter 3 yields an entire set \mathbf{U} of utility functions, which need not be related by a positive ratio transformation. For example, \mathbf{U} might contain $u(\cdot)$ and all positive multiples, and $v(\cdot)$ and all positive multiples, where $u(\cdot)$ and $v(\cdot)$ are not positive multiples of each other. The case in which \mathbf{U} consists of a single $u(\cdot)$ and all positive multiples is a limiting case – in which all individuals have complete and homogeneous (fully informed, fully rational, self-interested) preferences over life-histories and lotteries. This is *possible* – but my account does not insist on it. The account allows that individuals' preference over life histories and lotteries might be incomplete and/or heterogeneous.

Still, I will facilitate the presentation at various points in this chapter by discussing examples where \mathbf{U} does consist of a single utility function and all positive multiples thereof. This is done purely for expository purposes: nothing in my case for a prioritarian SWF will at any point hinge on the premise that utility is unique up to a positive ratio transformation. Conversely, because the case in which utility is uniquely measurable – where \mathbf{U} is a singleton – is not consistent with the best account of well-being, I will not employ examples of that sort.

Note that the utilitarian choice is to move to the suburb: the gain to the first child of living in the suburb rather than the city ($25c-20c = 5c$) exceeds the loss to the second child ($10c-9c$).³⁰ But any non-utilitarian will find plausible the thought that moving to the city is the better choice.

The thrust of Parfit's Lindley Lecture is to distinguish between two different kinds of non-utilitarian rationales for moving to the city. One such rationale, according to Parfit, is an "Egalitarian" view, which endorses what Parfit calls "The Principle of Equality": "It is in itself bad if some people are worse off."³¹ However precisely one measures the distribution of well-being within a given outcome, it seems intuitively very clear that the distribution of well-being in the city outcome ($20c, 10c$) is more equal than the distribution of well-being in the suburb outcome ($25c, 9c$).

The second possible rationale that Parfit offers for moving to the city is what he terms "The Priority View": "Benefitting people matters more the worse off these people are."³² The idea, here, is to compare the city and suburb outcomes by looking at each child's gain or loss in moving to the suburb and then *adjusting* each gain or loss to take account of the child's level of well-being, rather than simply summing the gains and losses in utilitarian fashion. As Parfit explains:

For Utilitarians, the moral importance of each benefit depends only on how great this benefit would be. For *Prioritarians*, it also depends on how well off the person is to whom this benefit comes. We should not give equal weight to equal benefits, whoever receives them. Benefits to the worse off should be given more weight.³³

One aspect of "prioritarianism," then, is that it gives greater weight to well-being changes affecting worse-off individuals. An analogy can be drawn, here, to the connection between money and well-being. It is very plausible that, as a matter of the best account of well-being, money has diminishing marginal utility. Both prioritarians and utilitarians recognize that connection. In addition, however, prioritarians believe that well-being itself has "diminishing marginal moral importance,"³⁴ as Parfit puts it.

A different aspect of prioritarianism, also stressed by Parfit in the Lindley Lecture, is that it is unconcerned with "relativities."

[O]n the Priority View, we do not believe in equality. We do not think it in itself bad, or unjust, that some people are worse off than others. This claim can be misunderstood. We do of course think it bad that some

³⁰ There are presumably other members of the population (at a minimum, the parent!), but they are assumed to be unaffected: each is just as well off regardless of where the family moves.

³¹ Parfit (1991; 84).

³² *Id.* at 101

³³ *Id.* at 101

³⁴ *Id.* at 105

people are worse off. But what is bad is not that these people are worse off than *others*. It is rather that they are worse off than *they* might have been.

Consider ... the central claim of the Priority View: benefits to the worse off matter more. ... On this view, if I am worse off than you, benefits to me are more important. Is this *because* I am worse off than you? In one sense, yes. But this has nothing to do with my relation to you.

It may help to use this analogy. People at higher altitudes find it harder to breathe. Is this because they are higher up than other people? In one sense, yes. But they would find it just as hard to breathe even if there were no other people who were lower down. In the same way, on the Priority View, benefits to the worse off matter more, but that is only because these people are at a lower *absolute* level. It is irrelevant that these people are worse off *than others*. Benefits to them would matter just as much even if there *were* no others who were better off.

The chief difference is, then, this: Egalitarians are concerned with *relativities*: with how each person's level compares with the level of other people. On the Priority View, we are concerned only with people's absolute levels.³⁵

It also bears mention that Parfit differentiates between what he terms the “telic,” i.e., outcome-oriented, and “deontic” variants of prioritarianism and egalitarianism. Given the consequentialist approach of this book, my focus will be on the “telic” versions.

The Lindley lectures, as mentioned, have generated a substantial literature.³⁶ One central point of confusion and contention in this literature is whether (telic) prioritarianism refers to a particular type of ordering of outcomes, or a particular *reason* or *justification* for an ordering. Indeed, there is confusion about what that distinction would amount to.

Within my set-up, the distinction is easy to formulate. An SWF is a mathematical rule, ranking two outcomes as a function of their associated utility vectors.³⁷ A justification for an SWF is a certain pattern of moral argumentation for one or another such rule. We can categorize the rules *and* categorize the patterns of argumentation for them – in both cases, doing so in order to advance our goal of identifying a particular SWF that we can accept in reflective equilibrium.

Like some other scholars, I find it useful to employ the term “prioritarianism” to identify a particular category of SWFs. Used in this manner, what does the term mean? The answer to *that* question is pretty easy. As mentioned, Parfit's seminal discussion points to two aspects of prioritarianism: first, that it gives more weight to well-being changes affecting worse-off individuals; and, second, that it is unconcerned with “relativities.” Understood as a way to delineate a subset of SWFs, the first aspect is naturally translated as the Pigou-Dalton axiom.

³⁵ *Id.* at 104.

³⁶ The contemporary philosophical literature on prioritarianism includes: Brighthouse & Swift (2006); Brock (___); Broome (___); Brown (2005); Crisp (2003); Fleurbaey (___); Hausman (___); Holtug (2007); Jensen (2003); Lumer (2005); Mason (2006); McKerlie (1994); Persson (2001); Peterson & Hansson (2005); Rabinowicz (2001); Tungodden (2003), as well Temkin's scholarship, see below, which engages prioritarianism at many junctures. See also Weirich (1983).

³⁷ More precisely, a utility set \mathbf{U} associates each pair of outcomes with a set of pairs of utility vectors; and an SWF is a rule for ranking the outcomes depending on this utility information. See above note 11.

How better to formulate the idea that well-being has “diminishing marginal moral weight?” Understood as a way to delineate a subset of SWFs, the second aspect is naturally translated as separability. How better to formulate the idea that an SWF is focused on where each individual is located on the scale of well-being, rather than on a comparison between individuals’ levels?

We have thus arrived at the conclusion that the term “prioritarian” is usefully attached to Pigou-Dalton respecting, separable SWFs. Indeed, the scholars who have associated prioritarianism with a particular type of SWF have generally reached either this conclusion, or the related idea that a prioritarian SWF takes the form $\sum_{i=1}^N g(u_i(x))$.³⁸ Nothing in Parfit’s basic insights rules out leximin³⁹, and for this reason (as well as others)⁴⁰ I will use “prioritarianism” to mean the entire family of Pigou-Dalton, separable SWFs, including but not limited to SWFs that satisfy a continuity requirement and use the $\sum_{i=1}^N g(u_i(x))$ formula.

Utilitarianism versus Distribution Sensitive SWFs

What would justify distribution-sensitive welfarism? Why not simply maximize overall well-being?

As I see it, non-utilitarian welfarism is most plausibly justified by considerations of *fairness*. For example, in Nagel’s two-child case, moving to the city is, intuitively, a *fairer* outcome than moving to the suburb – and therefore, perhaps, a morally preferable outcome, all things considered – even though the aggregate well-being of the two children would be greater in the suburb. Further, the two plausible conceptions of fairness that I will articulate, later in this chapter – albeit distinct in important ways -- can each be appealed to in support of this intuition.

The utilitarian may hold fast. Even with fairness theorized, she may not feel its tug.⁴¹ At this point, she might be asked: If considerations of fairness are irrelevant, what justifies *welfarism*? Remember that welfarism, as I have framed it here – consistent with the practice and theory of welfare economics – is *person-centered*. The well-being of sentient non-human

³⁸ For discussions of how to formally represent prioritarianism, see Broome (____); Brown (2005); Jensen (2003); Fleurbaey (____); Holtug (2007); Lumer (2005); Rabinowicz (2001); Tungodden (2003).

³⁹ Parfit does characterize leximin as too extreme, see Parfit (1991; 121), and I will ultimately agree; but the point is that nothing in the basic insight that we might end up preferring redistribution in virtue of a moral view that does not focus on relativities and gives greater moral weight to well-being changes affecting worse-off individuals rules out leximin.

⁴⁰ Classifying the leximin SWF as one type of prioritarian SWF will help to structure my analysis of different types of SWFs, in light of different conceptions of fairness. See below.

⁴¹ A different possibility for the utilitarian is to adopt an across-outcome conception of fairness, but argue that the measure of each person’s claim is simply his well-being difference. I believe that such a construal of the across-outcome conception -- which not only fails Pigou-Dalton, but *never* sees a pure well-being transfer as improving fairness – is very implausible. See below.

animals is ignored.⁴² (Consider that the Pareto principle, as framed by welfare economists, is almost always framed in terms of persons and their preferences, not animals; and that cost-benefit analysis, the dominant policy-analytic technique, is based on aggregating the money amounts that persons (not animals) are willing to pay or accept for their well-being changes.) If fairness is irrelevant to morality, why should morality focus on persons? What justifies person-centered utilitarianism rather than an inclusive utilitarianism that encompasses animal as well as human well-being?

The utilitarian may have an answer to this question. She may say: “Human well-being is qualitatively different from animal well-being. The concept of the ‘moral’ can be used to demarcate this difference. We have *moral* reason to maximize human well-being, and non-moral reason to maximize animal well-being.” This explanation for why morality should be both person-centered and utilitarian is not incoherent. But a stronger explanation for the person-centered cast of morality points to fairness – namely, that fairness norms must be such that their beneficiaries can be guided by these norms, and deliberate about these norms, and only persons are capable of being guided by and deliberating about norms. We do indeed have (non-moral) reason to be concerned with animal well-being, but ignoring animals’ interests is not *unfair* to them. Animals have no claim on us as a matter of distributive justice.⁴³

Much more can surely be said, here. But the basic idea that utilitarianism is problematic because it is unfair to persons – that utilitarianism “ignores the separateness of persons,” as Rawls puts it – is no doubt familiar to the reader, and I will not belabor it. Rather, the remainder of this section will survey a variety of less familiar considerations relevant to the choice between utilitarian and distribution-sensitive welfarism. “Granted,” the utilitarian might say, “the notion of fairness seems to make a *prima facie* case for distribution-sensitive welfarism, but don’t the following considerations tilt the other way?”

Harsanyi’s Impartial Observer

John Harsanyi presents two separate arguments for utilitarianism.⁴⁴ One, encapsulated in his impartial observer theorem, hinges on the idea that the moral ranking of outcomes depends on how individuals would self-interestedly rank them under a probabilistic veil of ignorance.

I will rework the argument somewhat, presenting it as follows. Imagine that individuals have self-interested extended preferences regarding life-histories and lotteries which are consistent with expected utility (EU) theory. And imagine further, as Harsanyi supposes, that

⁴² This is formally captured in my stipulation that any welfarist decision procedure should, at a minimum, generate a quasiordering of an outcome set which satisfies the Pareto principles – that it should rank as equally good a pair of outcomes in which the N human persons who exist are unaffected, and that it should rank as better an outcome in which at least one person is better off and all at least as well off, *regardless* of the impact of the outcomes on non-human animals.

⁴³ For a fuller discussion, see Chapter 1.

⁴⁴ For Harsanyi’s scholarship and critical reactions, see Chapter 3.

individuals have identical such preferences. (As already mentioned, I believe this supposition may fail; but let us grant Harsanyi the supposition for the sake of argument.) The result is that individuals' (identical) self-interested preferences over life histories and lotteries can be expectationally represented by a utility function, $u(\cdot)$.⁴⁵ EU theory ensures that $u(\cdot)$ is unique up to a positive affine transformation; if we add in comparisons to nonexistence, $u(\cdot)$ becomes unique up to a positive ratio transformation, and so we can represent the utility of any life-history $(x; i)$ as $cu(x; i)$, c any positive number.

Let us now add the further premise that individuals take an impartial or *moral* perspective on outcomes by viewing each outcome as an equiprobability lottery over its component life-histories, and determining which lottery is self-interestedly preferred. Outcome x is evaluated, morally, by seeing it as a $1/N$ probability of life-history $(x; 1)$; a $1/N$ probability of life-history $(x; 2)$; ... ; a $1/N$ probability of life-history $(x; N)$. Similarly, any other outcome y is evaluated, morally, by seeing it as a $1/N$ probability of life-history $(y; 1)$; a $1/N$ probability of life-history $(y; 2)$; ... ; a $1/N$ probability of life-history $(y; N)$.

If this further premise is correct, it follows that each individual weakly morally prefers⁴⁶ outcome x to outcome y just in case $(1/N) cu(x; 1) + (1/N) cu(x; 2) + \dots + (1/N) cu(x; N) \geq (1/N) cu(y; 1) + (1/N) cu(y; 2) + \dots + (1/N) cu(y; N)$. Multiplying both sides by N , we have that each individual weakly morally prefers x to y just in case the sum of the utilities of the component life-histories of x is at least as large as the sum of the utilities of the component life-histories of y . In short, it seems, each individual's moral preferences are utilitarian.

A standard objection to this argument is that there is no reason to assume that the utility number $cu(x; i)$ assigned to a given life history is the same as its well-being. However, for reasons discussed at length in Chapter 2, I believe that this objection fails. As I argued there, Harsanyi is very plausibly correct to use EU theory to construct a cardinal measure of the well-being associated with life histories, in more or less the manner he suggests.⁴⁷ In the case at hand, if we stipulate that individuals' preferences over life-history lotteries are identical, self-interested and also suitably idealized,⁴⁸ then $u(\cdot)$ and all ratio transformations thereof are indeed accurate measures of the well-being associated with life-histories. It is indeed the case that one life-history is better for well-being than another iff its $cu(\cdot)$ value is higher; and it is indeed the case

⁴⁵ To say that $u(\cdot)$ expectationally represents the lotteries is just to say that the lotteries are ranked according to the probabilistic expectation of $u(\cdot)$. This is the central idea of EU theory. See Chapter 3 for a discussion.

⁴⁶ To weakly prefer is to prefer or be indifferent.

⁴⁷ I stipulate that individuals' extended preferences must be idealized and self-interested, and I allow for heterogeneity in such preferences. But like Harsanyi I use EU theory to construct a utility function that expectationally represents each individual's extended preferences, and thus to arrive at a measure of the differences between life histories.

⁴⁸ In my view, it is critical to add these conditions, because a utility function that expectationally represents individuals' non-ideal or non-self-interested preferences over life histories will *not* necessarily be the same as $cu(\cdot)$, the utility function that represents the *well-being* ranking of histories and differences.

that differences between life-histories are accurately measured by differences between their $cu(\cdot)$ values.

My objection to Harsanyi's argument, here, is different. Why assume that morally evaluating outcomes just means viewing them as equiprobability lotteries over their component life-histories, and self-interestedly ranking those lotteries?⁴⁹ Why view *that* as the essence of moral impartiality? The welfarist who is sensitive to considerations of fairness can reply that the moral ranking of outcomes should not be understood in this manner. For example, she can argue (as I will below) that the moral ranking of outcomes should be constructed by viewing individuals as having claims between outcomes – claims which need not be linear in the individuals' well-being differences between outcomes.

I suppose Harsanyi could reply that we have strong pretheoretical intuitions in favor of the probabilistic-veil-of-ignorance construal of what it means to morally rank outcomes. Such an intuition, together with the procedure for assigning well-being to life-histories that Harsanyi employs (and that I more or less concur with), *could* push us to reject the prima facie moral case for distribution-sensitivity and to endorse utilitarianism. But does the reader have this intuition? Perhaps she has read Rawls. Whatever the attractiveness of Rawls' approach, it is not Harsanyi's. Rawls models the principles of justice as arising from choice behind a *nonprobabilistic* veil of ignorance – and it leads him, of course, not to utilitarianism, but to a maximin rule.

Measure-Theoretic Arguments; Invariance

The theoretical literature on SWFs has analyzed the linkages between various limitations on the measurability of well-being and the possible form of the SWF.⁵⁰ In general, the structure of the analysis runs as follows. It is stipulated that the SWF must provide a complete ranking of outcomes, if each outcome is associated with a single utility vector by a single utility function $u(\cdot)$. The question is then asked: If we require the ranking to be the same if we replace $u(\cdot)$ with some stipulated transformation, $\varphi(u(\cdot))$, which SWFs will satisfy this requirement? The idea is that $\varphi(u(\cdot))$ might contain all the information about well-being contained in $u(\cdot)$, and that – if so—the ranking of outcomes yielded by an attractive SWF should be invariant to whether we measure well-being using $u(\cdot)$ or $\varphi(u(\cdot))$.

For example, the following fairly striking result can be shown. Imagine that $\varphi(u(\cdot))$ is stipulated to be any positive affine transformation of $u(\cdot)$.⁵¹ The idea, here, would be that any positive affine transformation produces the very same ranking of life-histories and differences between them as $u(\cdot)$ itself, and thus contains all the information in $u(\cdot)$. If, in addition, we

⁴⁹ See Moreno-Ternero & Roemer (2008).

⁵⁰ See Boadway & Bruce (1984, chapter 5); Bossert & Weymark (2004); d'Aspremont & Gevers (2002); Mongin & d'Aspremont (1998).

⁵¹ In other words, the utility assigned by $u(\cdot)$ to each life-history is multiplied by a positive constant, and a constant is added.

demand that the SWF be impartial, Paretian, and separable, and be either a distribution-sensitive or utilitarian SWF, then (roughly) the SWF must either be leximin or utilitarianism.⁵² If one adds in a continuity requirement (or just rejects leximin as too implausible), the upshot is the utilitarian SWF.

But this invariance requirement is too strong. Although $u(\cdot)$ and any positive affine transformation thereof *do* yield the same information about the ranking of life-histories and their differences, they do not necessarily contain the same information about the *ratios* between life-histories. As I argued in Chapter 3, ratio information can be meaningfully constructed by assigning a utility of zero to nonexistence. For any utility function $u(\cdot)$ that assigns zero to nonexistence, $\varphi(u(\cdot))$ produces the very same ranking of life-histories and differences between life-histories as $u(\cdot)$, and contains the very same ratio information, if $\varphi(u(\cdot))$ is a positive multiple of $u(\cdot)$ ⁵³ -- but not necessarily if $\varphi(u(\cdot))$ is a positive affine transformation.

So a reasonable invariance requirement to impose on a SWF should be this: the ranking of outcomes according to a given utility function should be unchanged if we replace that function by a positive multiple. This will permit, not just leximin, and not just a utilitarian SWF, but also a prioritarian SWF – in particular, the Atkinsonian SWF.

Broome’s Measure Theoretic Argument

John Broome has developed a different kind of measure-theoretic argument. Rather than being an argument for utilitarianism, this argument suggests that prioritarianism and utilitarianism are not in fact distinct approaches to ordering outcomes.

The essence of Broome’s argument is this: “To give their theory meaning, prioritarians need a measure of a person’s wellbeing that is distinct from the value of her well-being. They may not be able to find one.”⁵⁴

Prioritarians agree with utilitarians that each life-history can be assigned a well-being number $u(x; i)$, or at least a set of such numbers. Prioritarians also agree with utilitarians that the *moral* value of an outcome is the sum of the moral values of the life-histories it contains. *Both* prioritarians and utilitarians assign a moral value to an outcome by using a general formula with the structure $w(x) = \sum_{i=1}^N v(x; i)$. Utilitarians, however, say that the moral value of a life-history is the very same as its well-being value. They equate $v(x; i)$ and $u(x; i)$. Prioritarians, by

⁵² More precisely, if we require that the ordering of utility vectors be Paretian, anonymous, separable, and satisfy a “minimal equity” condition which is designed to rule out leximax and is satisfied by all plausible SWFs (both utilitarianism and all distribution-sensitive SWFs), the only possibilities are leximin or “weakly utilitarian” SWFs, which are all SWFs (including but not limited to utilitarianism) that agree with utilitarianism in ranking utility vectors with different total amounts of utility. See, e.g., Bossert & Weymark (2004; 1157-58).

⁵³ In other words, for all life histories, $\varphi(u(x; i)) = cu(x; i)$, with c a positive constant.

⁵⁴ Broome (____; 4). See Broome (1991; 213-22); Rabinowicz (2001; 143-47); Jensen (____).

contrast, distinguish between the two. They say that $v(x; i) = g(u(x; i))$, with $g(\cdot)$ some strictly increasing and strictly concave function. But what procedure can the prioritarian use to attach these two different numbers to a life history: on the one hand, $u(x; i)$; on the other hand, $g(u(x; i))$?

Consider the following procedure. We *start* with the moral ordering of outcomes which, for simplicity, may be assumed to be complete. We *start* with the proposition that outcome x is morally better than outcome y , which is morally better than z , and so forth. We then identify a function for ascribing values to life histories, $v(\cdot)$ which “tracks” the ordering of outcomes. We identify a $v(\cdot)$ such that x is morally at least as good as y iff $\sum_{i=1}^N v(x; i) \geq \sum_{i=1}^N v(y; i)$.

This procedure will not enable us to ascribe two further values to a given $(x; i)$, namely $u(x; i)$ and $g(u(x; i))$. Imagine that, having identified $v(\cdot)$, we propose a particular $u(\cdot)$ function as the “right” function. This $u(\cdot)$ function is such that $v(x; i) = g(u(x; i))$, with $g(u(x; i)) = \sqrt{u(x; i)}$. It looks like we have arrived at prioritarianism, because setting $g(\cdot)$ as the square root function yields a strictly increasing and strictly concave $g(\cdot)$. But the very same set of v values assigned to life histories would be produced by defining a new utility function $u^*(x; i) = \sqrt{u(x; i)}$, and defining a new $g(\cdot)$ function $g^+(u^*(x; i)) = u^*(x; i)$. This new $g(\cdot)$ function is not strictly concave.

This line of questioning is, I take it, the essence of Broome’s skepticism about distinguishing between utilitarianism and prioritarianism. The response is that there are some procedures that *do* allow us to ascribe two different values to a life history, namely $u(x; i)$ and $g(u(x; i))$. Specifically, my discussion of the well-being quasiordering in the preceding chapter does *not* propose that we ascribe utility values to life-histories by reasoning *from* the moral ordering of outcomes. Rather, we measure the well-being associated with life-histories by determining the utility numbers that expectationally represent the self-interested preferences of fully-informed and rational individuals, considering different possible lotteries over life histories and comparisons of life-histories to nonexistence. So we have a basis quite independent of our judgments about the moral ordering of outcomes for constructing the set \mathbf{U} of utility functions.

More generally, any welfarist theory that contains an account of well-being which provides some basis for constructing \mathbf{U} that is independent of the moral ordering of outcomes, and instead uses these utilities as the inputs for that ordering, can resist Broome’s argument that prioritarianism and utilitarianism are indistinguishable.

Harsanyi’s Aggregation Theorem

Harsanyi offers two distinct arguments for utilitarianism: one based on the notion of an impartial observer, the other in his so-called “aggregation theorem.”

The basic idea of the aggregation theorem can be stated as follows: if the moral ranking of choices complies with expected utility theory, and if in addition the moral ranking of choices must satisfy the *ex ante* Pareto principles -- *ex ante* Pareto indifference and *ex ante* Pareto superiority -- that ranking must be utilitarian.

The validity of the aggregation theorem, as Harsanyi presents it, is controversial. I will not review that controversy. But, on Harsanyi's behalf, it should be pointed out that the following argument, roughly along the lines of Harsanyi's, is valid. Assume that utility is unique up to a positive ratio transformation. Using this utility information, some SWF produces a ranking of an outcome set. Assume, further, that the moral ranking of choices that might yield those outcomes is consistent with EU theory.⁵⁵ More specifically, imagine that the moral ranking of a hypothetical choice set, consisting of all possible lotteries over the outcomes, would be consistent with EU theory. Assume, finally, that the ranking of this hypothetical choice set satisfies the *ex ante* Pareto principles: If there are two choices, *a* and *b*, such that each individual's expected well-being with *a* is the same as her expected well-being with *b*, then the choices must be ranked as morally equally good. And if there are two choices, *a* and *b*, such that at least one person has greater expected well-being with *a*, and everyone's expected well-being is at least as great, then *a* must be ranked as the morally better choice.

It can be shown that any SWF that satisfies these requirements as well as the basic anonymity requirement must yield the very same ranking of the outcome set as the utilitarian SWF.⁵⁶

To see, in particular, why a continuous prioritarian SWF is ruled out by the assumptions (1) that the moral ranking of choices satisfies EU theory, and (2) that this ranking must comply

⁵⁵ It should be stressed that EU theory is being used here (as in Chapter 7) to guide the moral ranking of choices that would yield various probability distributions over outcomes, and not (as in Chapter 3) to guide the well-being ranking of choices that would yield various probability distributions over life histories.

⁵⁶ On different versions of the aggregation theorem, see Weymark (1991). Assume that U is unique up to a positive ratio transformation, i.e., consists of $u(\cdot)$ and all positive multiples. One version of the aggregation theorem shows that, if each individual i 's preferences over a lottery set are consistent with EU theory, thus can be expectationally represented by some utility function; and if social preferences over the lottery set are consistent with EU theory *and* satisfy *ex ante* Pareto indifference and superiority; *then* social preferences can be represented as an affine function of individual utilities, all with nonnegative weights. Translating this theorem to the context at hand: given the outcome set O , and a hypothetical set of all lotteries over O , each individual i 's well-being can be expectationally represented by $u_i(\cdot)$. The expected value of $u_i(\cdot)$ tracks the goodness of the lotteries for individual i 's well-being. So there is a moral utility function, which expectationally represents the moral ranking of the hypothetical lotteries, and is the sum of individual $u_i(\cdot)$ values, with nonnegative weights. In particular, the moral ranking of outcomes (degenerate lotteries) is simply the sum of individual $u_i(\cdot)$ values, with nonnegative weights.

Imagine, now that the outcome set has a pair of two outcomes which are permutations of each other. (Each individual in one outcome can be mapped onto one individual in the other outcome with the very same well-being level.) Anonymity requires that the SWF rank these two outcomes as equally good, which in general will not occur if the weights in the above paragraph are unequal. Thus (at least in the case of an outcome set with a sufficient number of pairs of outcomes that are permutations of each other, and at least in the case of utilities unique up to a positive ratio transformation), any anonymous SWF that ranks hypothetical lotteries consistent with EU theory and satisfies the *ex ante* Pareto principles must rank the outcome set exactly the same way as a utilitarian SWF. See Broome (1991), an influential work which uses a version of the aggregation theorem to argue for utilitarianism.

with ex ante Pareto indifference and ex ante Pareto superiority, consider the following example, a probabilistic variation on Nagel's two-child case. Imagine that a parent has choice of hiring two caregivers: Biased Bill or Evenhanded Earl. The parent is sure that Bill would focus his attention on one child, to the neglect of the other. The favored child would end up with utility $25c$, the neglected one with utility $9c$. The parent believes there is a 50% chance that Bill would favor either child. By contrast, the parent is sure that Earl would devote his attention evenhandedly to both kids, producing utility of $17c$ for each, or perhaps a little less.

Any decisionmaker who uses a continuous prioritarian SWF and ranks the choices consistent with EU theory must prefer hiring Earl to hiring Bill. And yet ex ante Pareto indifference and ex ante Pareto superiority require that hiring Bill is either equally good as hiring Earl (if Earl produces $17c$ for each child) or worse (if Earl produces less than $17c$), as the following tables illustrate.

Why the application of a continuous prioritarian SWF under uncertainty cannot satisfy both EU theory and the ex ante Pareto principles

	<i>Hiring Bill</i>		expected	<i>Hiring Earl</i>		expected
	<u>p=.5</u>	<u>p=.5</u>	<u>utility</u>	<u>p=.5</u>	<u>p=.5</u>	<u>utility</u>
First child	$25c$	$9c$	$17c$	$17c$	$17c$	$17c$
Second child	$9c$	$25c$	$17c$	$17c$	$17c$	$17c$

In the case represented by the tables above, hiring Bill produces an expected utility of $17c$ for each child, as does hiring Earl, so ex ante Pareto indifference requires that the two choices be ranked equally good. But any continuous prioritarian SWF will say that each possible outcome of hiring Bill (giving $25c$ to one child and $9c$ to the other) is worse than each possible outcome of hiring Earl ($17c$ for each child). The EU approach to applying an SWF under uncertainty implies (among other things) that, if each possible outcome of action a is ranked better than each possible outcome of action b by the SWF, a will be chosen. **Therefore** the EU approach conjoined with a continuous prioritarian SWF picks the option of Hiring Earl, in violation of ex ante Pareto indifference.

	<i>Hiring Bill</i>		expected	<i>Hiring Earl</i>		expected
	<u>p=.5</u>	<u>p=.5</u>	<u>utility</u>	<u>p=.5</u>	<u>p=.5</u>	<u>utility</u>
First child	$25c$	$9c$	$17c$	$(17-\epsilon)c$	$(17-\epsilon)c$	$(17-\epsilon)c$
Second child	$9c$	$25c$	$17c$	$(17-\epsilon)c$	$(17-\epsilon)c$	$(17-\epsilon)c$

In the case represented by *these* tables, hiring Bill produces an expected utility of $17c$ for each child, while hiring Earl produces an expected utility of $(17 - \epsilon)c$ for each child, so ex ante Pareto superiority requires that hiring Bill be preferred. But any continuous prioritarian SWF will say that the outcome of giving $(17 - \epsilon)c$ to each child is better than the outcome of giving $25c$ to one child and $9c$ to the other, for ϵ sufficiently small. **Therefore**, for ϵ sufficiently small, the EU approach conjoined with a continuous prioritarian SWF picks the option of hiring Bill, in violation of ex ante Pareto superiority.

My answer to this argument will be developed in Chapter 7. That chapter argues that there is no reason to insist that the moral ranking of choices satisfy conditions of ex ante Pareto indifference and superiority. This means (as I will argue in that chapter) that the moral ranking of an outcome set **O** produced by a prioritarian SWF should indeed be employed to generate a moral ranking of a choice set **A** in a manner consistent with EU theory – notwithstanding violations of the ex ante Pareto principles. It also means that the argument *against* prioritarianism, and in favor of utilitarianism, now on the table, should be rejected.

Distribution Sensitivity, Separability, and Two Conceptions of Fairness

For the remainder of the chapter, I assume that fairness argues in favor of some type of distribution-sensitive SWF, and that various possible utilitarian arguments, mooted in the prior section, have been defeated.

This section tries to establish the case for a distribution-sensitive SWF which is *separable*, as against non-separable distribution-sensitive SWFs. The section begins by introducing two different conceptions of fairness: a within-outcome conception (what Temkin terms “comparative fairness”) and an across-outcome conception, which I associate with the work of Thomas Nagel. As subsequent sections will show, the question of how fairness is best specified is relevant at many junctures in sorting between different types of distribution-sensitive SWFs. In particular, the question is key (I believe) to adjudicating the issue of separability.

Having introduced the two different conceptions of fairness, I suggest that the across-outcome conception is most plausibly seen to justify an SWF which satisfies separability as well as the Pigou-Dalton principle, while the within-outcome conception is most plausibly seen to justify an SWF which fails separability as well as the Pigou-Dalton principle.

This claim, standing alone, does not suggest that the within-outcome conception of fairness is any less attractive than the across-outcome conception. It simply clarifies how the two conceptions are most attractively specified. Proponents of prioritarianism need an independent argument *against* the within-outcome conception. I argue, here that the across-outcome conception fits better with welfarism.

If the argument succeeds, I will also have shown that a welfarist decision procedure is more attractively specified using a separable rather than non-separable SWF. And I will have

established an interesting ancillary result: namely, that while many distribution-sensitive SWFs *can* be expressed as the product of an inequality metric and overall well-being, this decomposition of the SWF is misleading. A welfarist moral decision procedure that uses the SWF format has no useful role for an inequality metric.

Two Conceptions of Fairness

What does it mean to treat persons fairly? The contemporary philosophical literature on equality suggests two different types of answers to this question. One type, pressed most vigorously by Larry Temkin, suggests that fairness is a matter of how individuals fare *relative to each other*.⁵⁷ This is what Temkin terms “comparative fairness”: the fairness of a situation is a matter of comparing each individual’s attainment with respect to welfare, or some other currency, with everyone else’s attainment (with due sensitivity for each individual’s responsibility for his attainment). “[C]oncern about equality is a portion of our concern about fairness that focuses on how people fare relative to others. So our concern for equality is not separable from our concern for a certain aspect of fairness; they are part and parcel of a single concern. Egalitarians in my sense generally believe that it is bad for some to be worse off than others through no fault or choice of their own.”⁵⁸

Temkin illustrates the idea with the following example:

This example concerns a fairly “typical” poor person in the United States, whom I shall call “Ruth.” Ruth is not wretched, but she is a single parent of four, works at two jobs, drives an old car, worries how she will meet the payments on her two bedroom apartment, and has no idea how her children will afford college on her \$20,000 income. Many are deeply moved by the plight of people like Ruth in a land where *so* many others live in half million dollar homes, own fancy new cars, send their children to private schools, take expensive vacations, and have household incomes well over \$100,000.

Is it not clear that the extent to which many are moved by Ruth’s situation is heavily influenced not merely by how she fares in *absolute* terms, but by how she fares *relative to other members of her extraordinarily well-off society*? After all, we may suppose, at least Ruth has a roof over her head, indoor plumbing, a telephone, a TV, a car. Moreover, she is not living in a war-torn country, or ruled by a dictator, and she need not fear smallpox, tuberculosis, malaria, or diphtheria. She drinks safe water, eats three meals daily, and has a reasonably long life-expectancy. In short, without romanticizing the plight of America’s poor, it seems that for most of human history, someone as well off as Ruth would be amongst the very best off. Moreover, importantly, I think Ruth must probably be counted amongst the world’s fortunate, even taking full account of the genuinely bad effects of being poor in a rich society. ...

I suspect, then, that if the world did not include others who were even better off, so that Ruth was actually better off than *everyone* else, we would not be *nearly* as concerned to improve her situation as we now are, and that this would be so even on the assumption that the net changes in Ruth’s life balanced out,

⁵⁷ Temkin’s scholarship on equality, centered on the concept of comparative fairness, includes Temkin (1991; 2000; 2003a, 2003b, 2003c, 2003d, 2003e).

⁵⁸ Temkin (2003b; 62).

so that her absolute level in that situation would be *exactly* the same as it is now. Surely, our attitude towards America's poor is deeply shaped by the presence of so many other who are *so* much better off.⁵⁹

A different conception of fairness is suggested by Thomas Nagel's work on equality in his 1977 Tanner Lecture and then in his book *Equality and Partiality*.⁶⁰ Nagel's Tanner Lecture is a precursor to Parfit's presentation of prioritarianism in Parfit's Lindley Lecture (and indeed, as we have seen, Parfit begins the Lindley Lecture by referencing Nagel's work and discussing the two-child example that Nagel had provided). Nagel, like Parfit, presents the principle of giving greater weight to well-being changes affecting worse-off individuals as one possible non-utilitarian moral principle, and indeed employs the term "priority" to denote the moral view in which this principle figures. Unlike Parfit and much of the subsequent literature, Nagel uses the term "egalitarian" to denote that moral view. He writes: "The essential feature of an egalitarian priority system is that it counts improvements to the welfare of the worse off as more urgent than improvements to the welfare of the better off."⁶¹ But this semantic difference should not obscure the fact that Nagel does articulate and indeed endorse the principle that well-being has diminishing marginal moral weight – a principle that, in Parfit's work and the subsequent philosophical literature, becomes a central aspect of prioritarianism.

What is especially interesting, for our purposes, is Nagel's picture of the moral theory in which the priority principle figures. In the Tanner Lecture, Nagel suggests that any plausible moral theory will take seriously the separateness of persons – in particular, as Nagel sees it, the fact that each person has a distinctive perspective or "point of view" – and that any plausible moral theory will give equal weight to each person's point of view. In other words, on any plausible moral theory, morality is a matter of identifying each person's *claims* for or against different outcomes or actions, and then adjudicating between conflicting claims: "The units about which [the moral] problem arises are individual persons, individual human lives. Each of them has a claim to consideration."⁶² The particular moral theory that embraces the priority principle emerges by seeing individuals as having claims for or against *outcomes*, which depend on each individual's lifetime well-being, or how individuals fare with respect to particular aspects of lifetime well-being; and which are sensitive to well-being levels as well as changes. In the Tanner Lecture, Nagel writes:

[An egalitarian priority system] resembles utilitarianism formally, in being applied first to the assessment of outcomes rather than of actions. But it does not combine all points of view by a majoritarian method. Instead, it establishes an order of priority among needs and gives preference to the most urgent, regardless of numbers. In that respect it is closer to rights theory.....

... One problem in the development of this idea is the definition of the order of priority: whether a single, objective standard of urgency should be used in construing the claims of each person, or whether his interests should be ranked at his own estimation of their relative importance. In addition to the question of

⁵⁹*Id.* at 71.

⁶⁰ Nagel (1977; 1991, 63-74).

⁶¹ (Nagel 1977, 118).

⁶² *Id.* at 111.

objectivity, there is a question of scale. Because moral equality is equality between persons, the individual interests to be ranked cannot be momentary preferences, desires and experiences. They must be aspects of the individual's life taken as a whole: health, nourishment, freedom, work, education, self-respect, affection, pleasure.

But let me leave these questions aside. The essential feature of an egalitarian priority system is that it counts improvements to the welfare of the worse off as more urgent than improvements to the welfare of the better off. These other questions must be answered to decide who is worse off and who is better off, and how much, but what makes a system egalitarian is the priority it gives to the claims of those whose overall life prospects put them at the bottom, irrespective of numbers or of overall utility. Each individual with a more urgent claim has priority, in the simplest version of such a view, over each individual with a less urgent claim. The moral equality of egalitarianism consists in taking into account the interests of each person, subject to the same system of priorities of urgency, in determining what would be best overall.⁶³

Nagel then uses these ideas to discuss the two-child case, explaining: "It is more urgent to benefit the second child, even though the benefit we can give him is less than the benefit we can give the first child.... An improvement in his situation is more important than an equal or somewhat greater improvement in the situation of the first child."⁶⁴

In his subsequent book, *Equality and Partiality*, Nagel again sets forth this idea of a moral theory -- flowing from the separateness of persons and of each person's perspective -- which ranks outcomes by considering individuals' claims between them, giving priority to the worse off.

The impartial attitude is, I believe, strongly egalitarian both in itself and in its implications. As I have said, it comes from our capacity to take up a point of view which abstracts from who we are, but which appreciates fully and takes to heart the value of every person's life and welfare. We put ourselves in each person's shoes and take as our preliminary guide to the value we assign to what happens to him the value which it has from his point of view. This gives to each person's well-being very great importance ...

... The result is an enormous set of values deriving from individual lives, without as yet a method of combining them or weighing them against one another when they conflict, as they will in the real world. The question whether impartiality is egalitarian in itself is the question whether the correct method of combination will include a built-in bias in favor of equality, over and above the equality of importance that everyone's life has in the initial set of values to be combined.

Each if impartiality were not in this sense egalitarian in itself, it would be egalitarian in its distributive consequences because of the familiar fact of diminishing marginal utility....

But I believe that impartiality is also egalitarian in itself ... What it means is that impartiality generates a greater interest in benefiting the worse off than in benefiting the better off -- a kind of priority to the former....

....[This] does not rule out all ranking of alternatives involving different persons, nor does it mean that benefitting more people is not in itself preferable to benefitting fewer. But it does introduce a significant element of non-aggregative, pairwise comparison between the persons affected by any choice or policy. ...

⁶³ *Id.* at 116-18.

⁶⁴ *Id.* at 124.

The claims on our impartial concern of an individual who is badly off present themselves as having priority over *each* individual who is better off: as being ahead in the queue, so to speak.⁶⁵

Although Nagel does not use the term “fairness,” I believe that the “egalitarian priority system” he describes can be seen as a plausible conception of fairness.

Moreover, I believe that this Nagelian view of fairness, as well as the contrasting view suggested by Temkin, can both be adapted to play a role within welfarism. In adapting Nagel’s and Temkin’s views in this manner, I do not suggest that either author endorses “welfarism” in the strict sense of that term as used throughout the book – namely, a moral view that lacks agent-relative side constraints or options, and derives moral guidance for choice from an impartial ranking of outcomes, one that is wholly focused on human well-being. However, the work of each author points to a distinctive way to formulate a welfarist conception of fairness, which might then function to justify the choice of an SWF.⁶⁶ I will formulate these two conceptions as follows.

A Welfarist Conception of Fairness: The Within-Outcome Version

Whether outcome x is more or less fair than outcome y depends on the distribution of well-being in outcome x , and the distribution of well-being in outcome y . At a minimum, if the distribution of well-being in x is perfectly equal, and the distribution of well-being in y is unequal, then x is a more fair outcome than y .

A Welfarist Conception of Fairness: The Across-Outcome Version

Whether outcome x is more or less fair than outcome y depends on individuals’ claims for or against each outcome. A claim is a relation between an individual and a pair of outcomes, with four possible valences: given any pair of outcomes, x and y , an individual has a claim in favor of x , a claim in favor of y , no claim either way, or an incomparable claim. There is a basic connection between the valence of an individual’s claim and the valence of her well-being: If an individual is better off in x than y , she has a claim in favor of x ; if an individual is equally well off in both outcomes, then she has no claim either way; if she is incomparably well off, then she has an incomparable claim. At a minimum, any rule for ranking outcomes in light of individual’s claims must rank two outcomes as equally fair if no individuals have a claim either way; and must rank one outcome as fairer than another if some individuals have a claim in its favor and all other individuals have no claim either way. Where these conditions are not satisfied – where individuals’ claims are conflicting – the ranking of the outcomes should be a function of

⁶⁵ Nagel (1991; 64-68).

⁶⁶ Further, Temkin does suggest that comparative fairness can be used to rank outcomes, and specifically that we can do so by focusing on the distribution of well-being. See Temkin (1991; 10-12). Similarly, Nagel sees his “egalitarian priority” system as focused on outcomes, and sees each individual’s well-being as a major determinant of his claim for or against an outcome. (Nagel 1977; 1991).

some fair rule for measuring the strength of individuals' claims and adjudicating between them.

Both the “within-outcome” conception and the “across-outcome” conception of fairness involve comparisons of different individuals – how could “fairness” be otherwise? – but they differ, as it were, in the *modal* structure of such comparisons. On the within-outcome conception, how individuals would fare relative to others, were one outcome to occur (x), and how individuals would fare relative to others, were another outcome to occur (y), is what determines whether one or the other outcome is more fair. On the across-outcome conception, instead, we determine, for each individual, whether she has a claim to have the world turn out one way (x) rather than another (y), and then compare these inter-world claims to determine which outcome is more fair.

Both conceptions are echoed in ordinary moral thinking. Sometimes, we do think morally about a choice by asking whether the upshot of the choice would be to leave some person unfairly better off than others. But sometimes our ordinary moral thinking takes a different track. We try to determine which individuals have an *interest* in the choice, and we make the choice by ascertaining how strong the interests of the affected parties are, ignoring individuals who are *unaffected*. To say that individual i is unaffected by the choice between a and b is, more or less, to say that his well-being would be the same regardless of which is chosen. To ask whether i has a stronger interest in the choice than j is *not* to inquire into the comparative well-being of i and j given the upshot of a , or the comparative well-being of i and j given the upshot of b . Rather, it is, more or less, to determine whether i would be better off in one world (the upshot of a) or another (the upshot of b); to do the same for j ; and to determine how strong a moral interest each individual has in favor of one or the other outcome in virtue of his well-being difference between them.

Various objections might be leveled against my attempt to adapt Temkin's and Nagel's ideas in this manner, so as to function as arguments within welfarism. Two bear particular mention. One is that it is silly to try to merge welfarism and fairness. Fairness (whether in the within-outcome or across-outcome sense) should be sensitive not merely to the well-being levels that individuals realize in different outcomes, but also to whether they are responsible for those levels.⁶⁷

To this important objection I respond that, as discussed in Chapter 1, I see the SWF approach as a first approximation to a policy-evaluation approach which does take account of individual responsibility. It provides a framework that, I hope, can ultimately be refined in this manner. We can think of an SWF as an *approximate* tool for ranking a given outcome set \mathbf{O} , where some of the outcomes are such that some individuals are (partly or wholly) responsible for being badly off in those outcomes; and as an *exact* tool for ranking a outcome set \mathbf{O} , where each

⁶⁷See, e.g., Temkin (1991; 12-18; 2003b; 62); Nagel (1991, 71).

individual's well-being in every outcome is not her responsibility. In other words, for each x , and for each individual i in the population, any shortfall between i 's well-being level in x and a higher level, were outcome x to obtain, would be a matter of "pure luck." That shortfall is not properly seen as attributable, in any way, to i 's poor choices and, thus, is not her (sole or partial) responsibility. Under such conditions, ranking the outcomes in \mathbf{O} by using an SWF in the fashion discussed in this book – an SWF applied to vectors of utilities representing individuals' realized well-being levels – should be seen as fully justifiable even by the responsibility-sensitive welfarist.⁶⁸

To be sure, such an outcome set is an idealization. The outcome sets that actual decisionmakers confront will not have responsibility "washed out" in this manner. Nor will I take a stab at specifying "responsibility" and thereby specifying the conditions under which individuals lack responsibility for well-being shortfalls – except to assume that such conditions are possible. With that assumption in hand, the idealization can be used to think about the appropriate form of an SWF – the task undertaken in this chapter. In thinking about the across-outcome and within-outcome conceptions of fairness, and the SWFs they justify, we should assume that in all outcomes everyone's well-being level is a matter of pure luck. This is no more problematic than the use of many other sorts of hypothetical cases in philosophical argumentation --- nor, as far I can see, does it in any way stack the deck in favor of prioritarianism.

A second objection focuses on my construal of across-outcome fairness. Why insist that the valence of an individual's claim depends on her well-being valence? Why might not an individual have a claim in favor of x over y even though she is equally well off in both outcomes? Why must she be seen as having a claim in favor of x (as opposed to a claim in favor of y , no claim either way, or an incomparable claim), if she is better off in x ? For example, mightn't an individual have non-self-interested goals -- what Sen terms "agency goals" – that revalence her claims in these ways?

Suffice it to say that the across-outcome conception cannot function to justify the ordering of outcomes, within welfarism, if the link between an individual's well-being valence and claim valence is relaxed in this manner. If an individual can have a claim in favor of x despite being equally well off in x and y , the basic welfarist principle of Pareto indifference is threatened. If an individual can have a claim in favor of y despite being worse off there, Pareto superiority is threatened.

Before I discuss which SWFs these two conceptions of fairness best justify, several points bear noting. First, each conception is stated in a very generic way. With respect to the within-outcome conception: I assume that anyone who thinks of fairness as a matter of the distribution of well-being within each outcome will, at a minimum, endorse the principle that a

⁶⁸ Temkin adopts a similar approach, see (1991; 12).

perfectly equal distribution of well-being is fairer than an unequal distribution (holding fixed responsibility). This is consistent with a very large class of metrics for assessing the distribution of well-being within an outcome: not only all the standard inequality metrics suggested by economists, but others as well.

Similarly, I have framed the across-outcome conception so as to incorporate the basic ideas that an individual's claim valence depends on his well-being valence; that an outcome is fairer if some claims point in its favor and none against; and that two outcomes with no claims pointing either way are equally fair. But I have left open questions concerning how to rank outcomes where claims conflict: how to measure the *strength* of individuals' claims and how to *adjudicate* between an array of conflicting claims of various strengths.

My aim is to crystallize the basic idea that the fairness ranking of outcomes might be framed in a manner that involves intraworld comparisons, or in a manner that does not -- and thus to establish a plateau for further argumentation -- by setting forth two conceptions of fairness, each with *certain* entailments; but, also, not to build too much into the two conceptions, so as not to be accused of using definitional fiat to screen out views that are worthy of substantive argument.

A final point concerns the connection between fairness and the all-things-considered ranking of outcomes. Any welfarist decision procedure (as I present that idea in this book) will arrive at an all-things-considered ranking of an outcome set \mathbf{O} : a ranking that, at least, is a quasiordering and satisfies Pareto indifference and superiority. On the within-outcome conception of fairness, the fact that one outcome is more or less fair than another *cannot* be seen as the sole factor that determines the all-things-considered ranking. Why? The within-outcome conception will, necessarily, count an outcome x with a perfectly equal distribution of well-being as more fair than a Pareto-superior outcome with an unequal distribution y .⁶⁹ Therefore, the within-outcome conception must see the all-things-considered ranking of outcomes as justified by combining fairness consideration with certain other considerations.

In particular, I will assume, the within-outcome conception must see the all-things-considered ranking of outcomes as justified by the combination of fairness and *overall well-being*.⁷⁰

⁶⁹ This is not artificial. If the unfairness of an outcome is a matter of individuals doing better than others in that outcome (modulo responsibility), then what could be fairer than an outcome in which no one does better than anyone else? Indeed, Temkin agrees with this, as do all standard inequality metrics, which have the equality-as-a-lower-bound property. See Chapter 2.

⁷⁰ This is the standard understanding of both economists and philosophers about how a concern for equality can function as a component of a welfarist, Pareto-respecting moral view. In principle, there might be some factor, other than overall well-being, which – along with a measure of within-outcome fairness – produces a Pareto-respecting ordering of outcomes. But there is nothing in either the economics literature, or philosophy, that actually argues in favor of such a factor.

By contrast, the proponent of the across-outcome conception can see fairness, alone, as justifying the moral ranking of outcomes.⁷¹ Because of the connection between the valences of an individual's well-being and her claims, the ranking of outcomes justified by the across-outcome conception will satisfy Pareto superiority as well as Pareto indifference. This is not to say that the across-outcome conception *can't* be combined with overall well-being or other considerations. But it *needn't* be -- and indeed, for reasons I will suggest below, the simplest and most elegant version of welfarism does not combine across-outcome fairness with other factors.

Which SWFs do they justify?

Which type of SWF does each conception of fairness argue for? If we flesh out each conception in the most attractive and plausible manner, which SWF are we led to?

Start with the across-outcome conception. First, I suggest that this conception of fairness argues for an SWF that satisfies the Pigou-Dalton principle. Because I have outlined the view in a generic way, the Pigou-Dalton principle is not an *entailment*. For example, imagine that the strength of each person's claim between a given pair of outcomes x and y , given some utility function $u(\cdot)$, is measurable by a single number; its measure is just her utility difference between the outcomes; and the fairer outcome is determined by summing these measures, across persons, and seeing how the sum turns out across utility functions.⁷² Such an approach would appropriately link an individual's well-being and claim valence; and it would rank outcomes as more or less fair consistent with the basic requirements of the across-outcome conception; yet it would lead, not to an SWF that satisfies the Pigou-Dalton principle, but rather to a utilitarian SWF.

But I suggest that a utilitarian SWF, or any other SWF that fails the Pigou-Dalton principle, sits uncomfortably with the across-outcome conception. Imagine that individual i 's well-being level is lower than individual j 's in both outcomes under comparison; that outcome y is better for i , while outcome x is better for j ; that the well-being difference is exactly the same in both cases; and that everyone else is equally well off in both outcomes. In this situation, isn't i 's claim in favor of y stronger than j 's claim in favor of x ? What individual i stands to gain from outcome y rather than x obtaining is exactly what individual j stands to lose: the well-being differences are the same. That, alone, would suggest that the two individuals' claims have equal strength. But now add the fact that individual i is worse off in both outcomes than individual j

⁷¹ See Parfit (1991).

⁷² In other words, for a given pair of outcomes, x and y , and a utility function $u(\cdot)$, there is an N -entry claim vector, where the i th entry is the measure of individual i 's claim in favor of x , according to $u(\cdot)$, and equals $u(x; i) - u(y; i)$. A negative measure means that she has a claim in favor of y , according to $u(\cdot)$. For each $u(\cdot)$, sum the entries in this vector. The approach now on the table says that x is at least as fair an outcome as y iff this sum is nonnegative for all $u(\cdot)$ belonging to \mathbf{U} ; and that y is at least as fair as x iff this sum is nonpositive for all $u(\cdot)$ belonging to \mathbf{U} . (One outcome will be fairer than the other iff it is at least as fair and the other is not at least as fair as it; equally fair if each is at least as fair as the other; and incomparably fair if none of these hold true.) An individual has a claim in favor of x , full stop, if her claim is nonnegative for all $u(\cdot)$ and positive for some; a claim in favor of y , full stop, if her claim is nonpositive for all $u(\cdot)$ and negative for some; and no claim either way if her claim is zero for all $u(\cdot)$

(and remember that we are thinking about outcomes in which anyone's shortfall between her well-being and a higher level is not her responsibility). Individual i asks to be moved a given distance on a scale of human flourishing, from one level to another. Individual j asks to be moved the very same distance, from a third, even higher, level to a level that is yet more elevated. Doesn't individual i have a stronger claim than individual j ?

This line of thinking is at the core of Nagel's two articles, mentioned earlier. It is what leads him from the separateness of persons, and the idea of each person having a claim to being better off, to the conclusion that we should give priority to worse-off individuals.

Second, I suggest that the across-outcome conception of fairness argues for an SWF that satisfies the separability axiom.

Again, the across-outcome conception (even supplemented by the Pigou-Dalton principle) does not *entail* separability. In general, the SWF approach works as follows. Given two outcomes, x and y , and a set \mathbf{U} of utility functions that measures the well-being associated with life-histories, each $u(\cdot)$ belonging to \mathbf{U} produces a pair of utility vectors: the list of individual utilities in x according to that utility function, and the list of individual utilities in y according to that utility function. And the ranking of x and y , by the SWF, is then a function of the set of pairs of such vectors. An SWF satisfies the *separability* principle iff its ranking of the two outcomes depends solely on the set of pairs of *trimmed down* vectors: vectors that give the utility (according to each member of \mathbf{U}) only of affected individuals, i.e., those who are not equally well off in both outcomes.

To see how the across-outcome conception of fairness might yield an SWF that fails separability, considering the following "rank-weighted" approach to measuring and adjudicating between claims. Roughly speaking, the approach works as follows. (See the margin for a more precise treatment.) Given two outcomes, x and y , and a utility function, the strength of each individual's claim in favor of x or y is measurable by a single number: the difference between his rank-weighted utility in x and his rank-weighted utility in y . The fairer outcome, now, is determined by summing *these* measures, across persons, and seeing how the sums turn out across all utility functions.⁷³

⁷³ The rank-weighted approach uses a set of N weighting factors, a_1, \dots, a_N , which are strictly decreasing. Given outcomes x and y , and a utility function $u(\cdot)$, determine each person's utility rank in x , with 1 for the smallest utility, 2 for the next-smallest, and so forth, splitting ties arbitrarily (For example, if $u(x) = (4, 7, 2, 7)$, then the first person has rank 2, the second person can be given rank 3 or 4, the third person has rank 1, and the fourth person has rank 4 if the third has 3, otherwise 3.) Do the same for y . We can then produce an N entry claim vector, with each entry the measure of that individual's claim in favor of x according to $u(\cdot)$, with negative entries meaning a claim in favor of y according to $u(\cdot)$, calculated as follows: (1) In the case where there are no "rank switches," i.e., each individual's rank in x is the same as her rank in y , individual i 's claim in favor of x is $u(x; i)a_{r(x,i,u)} - u(y; i)a_{r(x,i,u)}$, where $r(x,i,u)$ is i 's rank in x according to $u(\cdot)$. This is the simple method for calibrating claims stated in the text. (2) Where there *are* rank switches, measuring claims this way can violate the across-outcome conception: it can assign unaffected individuals a nonzero claim, or assign individuals who are better off in x a negative claim. So we can calculate claims a different way. Calculate the total sum of rank-weighted utility for x , and subtract the total

Such an approach satisfies the basic dictates of the across-outcome conception (as well as satisfying the Pigou-Dalton principle); and yet it yields an ordering of outcomes that *fails* separability, as the following table illustrates.

How a rank-weighted SWF violates separability and makes the rank of unaffected individuals affect the strength of affected individuals' claims

	Utility Rank (from lowest to highest)			
	<u>First</u>	<u>Second</u>	<u>Third</u>	Rank-weighted sum
Outcome x	$20c$	$50c$	$200c$	$3(20c)+2(50c)+1(200c)=360c$
Outcome y	$20c$	$60c$	$175c$	$3(20c)+2(60c)+1(175c)=355c$
Strength of claim	—	$2(10c)$	$1(25c)$	
Outcome x^*	$50c$	$100c$	$200c$	$3(50c) + 2(100c) + 1(200c)=550c$
Outcome y^*	$60c$	$100c$	$175c$	$3(60c)+2(100c) + 1(175c) = 555c$
Strength of claim	$3(10c)$	—	$1(25c)$	

Explanation: Outcome $x = (20c, 50c, 200c)$ and outcome $y = (20c, 60c, 175c)$, while outcome $x^* = (100c, 50c, 200c)$ and outcome $y^* = (100c, 60c, 175c)$, with the entries in the vectors representing the utilities of individuals 1, 2, and 3. Consider ranking the x - y pair and the x^* - y^* pair using a rank weighted rule, with the simplest weighting scheme such that a weight of 3 is assigned to the lowest utility, 2 the second lowest, and 1 the highest. The calculations are given above, showing that x is ranked over y but y^* over x^* , in violation of separability.

Individual 1 is unaffected in each pair. So he has no claim as between x and y , or as between x^* and y^* . However, because he has the lowest utility rank in the x - y pair, and the second-lowest utility in the x^* - y^* pair, the strength of individual 2's claim changes between the pairs (from a strength $20c$ in the x - y case to a strength of $30c$ in the x^* - y^* case). Thus the sum of the claims of the two affected individuals (2 and 3) points in favor of x over y but in favor of y^* over x^* .

Why exactly does the rank-weighted approach to measuring and adjudicating between claims violate separability? This approach *does* assign unaffected persons a zero claim (by “unaffected,” I just mean persons who are equally well off in the two outcomes under

sum of rank-weighted utility for y . Call this amount ΔS . If $u(x; i) = u(y; i)$, assign i a zero claim. For all other individuals, assign them claims arbitrarily, so that the total sum of claims is ΔS and so that individuals with greater utility in x than y are assigned a positive claim, otherwise a negative claim. Because the rank-weighted SWF satisfies Pareto superiority, it should always be possible to do this.

We then say that outcome x is at least as fair as outcome y iff the sum of the entries of the claim vector is nonnegative, for all $u(\cdot)$ belonging to \mathbf{U} ; and y is at least as fair as x iff the sum is nonpositive, for all $u(\cdot)$. As with the “claim” translation of the utilitarian SWF, we can say that: An individual has a claim in favor of x , full stop, if her claim is nonnegative for all $u(\cdot)$ and positive for some; a claim in favor of y , full stop, if her claim is nonpositive for all $u(\cdot)$ and negative for some; and no claim either way if her claim is zero for all $u(\cdot)$.

Calibrating claims in this manner satisfies the basic requirements of the across-outcome conception. Moreover, it produces exactly the same ranking of outcomes as the rank-weighted SWF, discussed earlier. All that has been done here is to translate that SWF into the language of claims.

comparison), as well as assigning each person who is better off in one outcome a positive claim, and assigning each person who is incomparably well off an incomparable claim. So the valence of each individual's claims is appropriately related to the valence of her well-being. But the rank-weighted approach to measuring the *strength* of affected persons' claims takes account of the well-being levels of the unaffected persons in the two outcomes.

However, I believe that the across-outcome conception is most plausibly seen to justify a SWF that satisfies the separability requirement. We start with the idea that the fairness ranking of outcomes is a function of individual claims, with unaffected individuals having a zero claim (a stipulation which is both intuitively very plausible, and required to satisfy Pareto indifference). Fairness, on this view, requires that unaffected individuals have no *direct* say in the ranking of the outcomes. But by departing from separability, we allow them to have an *indirect* say in the ranking of outcomes. Why should they have this role? If it's unfair to see them as having an interest, why is it fair to allow their well-being to change our assessment of the balance of claims of those who *do* have an interest?

Do not be distracted by the thought that an individual's well-being in a given outcome may well be partly determined by her relation to other persons. For example, it is possible that individual *i* is worse off in outcome *x* than in outcome *y*, even though her non-relational attributes are the same in both outcomes, by virtue of the fact that there are many people in *x* who have more income than her, while there are few people in *y* who do. It is even possible that an individual's well-being in an outcome is partly determined by the pattern of *well-being* in that outcome. Any such relational determinants of well-being that may obtain should be *reflected* in the set **U** of utility numbers. If that set really does accurately represent the well-being ranking of life-histories, and well-being differences between life-histories, then it will be fully sensitive to the impact of the pattern of income, health, visible consumption, or even well-being, in a given outcome, on each individual's well-being in that outcome.

The question now on the table is different. We are assuming a set **U** that does accurately reflect well-being (whether constructed in the manner outlined in Chapter 3, or in some other manner). In particular, the level of each affected person's well-being in *x* and *y*, and the difference between these levels – taking full account of the potential effect of relational facts on individual well-being – is accurately represented by the utility functions in **U**. Why should the well-being levels of individuals who are genuinely unaffected (taking full account of relational facts) have a role at the stage of measuring and adjudicating between the claims of affected individuals – a role in determining how to move *from* the array of utility information capturing the well-being of those individuals, to an assessment of which outcome is most fair?

Turn, now, to the within-outcome conception of fairness. Here, I will argue that the conception is most plausibly fleshed out to justify an SWF that fails separability and fails the Pigou-Dalton condition.

Once more, this is not a matter of conceptual entailment. If we think of the fairness ranking of outcomes as determined by the inequality of well-being within outcomes, and the moral ranking of outcomes as determined by fairness (in this sense) plus overall well-being, it is *possible* for the ranking of outcomes to satisfy separability, or Pigou-Dalton, or both.

This point, indeed, is the font of much confusion in the contemporary literature on prioritarianism.⁷⁴ An important mathematical result, regarding inequality metrics, is the following. Imagine that $w(\cdot)$ is a continuous function which ranks utility vectors in a manner that satisfies Pareto superiority and anonymity, and prefers a perfectly equal distribution of well-being. Then we can use $w(\cdot)$ to construct a minimal inequality metric, $I^w(\cdot)$, such that the ranking of utility vectors achieved by $w(\cdot)$ is exactly the same as multiplying the total utility in a given vector times the degree of equality in that vector, as measured by $1 - I^w(\cdot)$.⁷⁵

This means that any SWF of the form:

⁷⁴ The literature often characterizes prioritarianism as a view that places no intrinsic value on equality; but it is puzzling what exactly this means, since many kinds of non-utilitarian orderings of outcomes, *including* orderings seen as “prioritarian,” can be represented as the result of a joint concern for overall well-being and the equality of well-being. See Brown (2005, chapter 3); Fleurbaey (___); Jensen (2003).

⁷⁵ By a “minimal inequality metric,” I mean a metric that assign a lower inequality score to a utility vector in which everyone’s utility is equal than to a vector in which utilities are unequal, and which is anonymous (assigns the same score to utility vectors that are permutations of each other), but may or may not satisfy the Pigou-Dalton principle, i.e., see a pure transfer of utility from a higher-utility to a lower-utility individual, without switching ranks, as decreasing the inequality score. (See Chapter 2, discussing standard inequality metrics, which satisfy equality-as-a-lower-bound, anonymity, *and* Pigou-Dalton.)

To produce a minimal inequality metric from a Paretian, anonymous, continuous $w(\cdot)$ with a preference for a perfectly equal distribution of utility, do the following. (See Blackorby et al. 2003, 68-128, for a discussion of how to construct an equality metric from an SWF.) Consider the “line of perfect equality,” i.e., all vectors in which utilities are equal. Let us use the abbreviation “ Σ ” to mean the total utility in a utility vector. And, to simplify presentation, I’ll use letters such u , v , and z to refer to utility vectors (omitting reference to the underlying outcomes that are mapped by utility functions onto one or another vector.) In the case of a Paretian, continuous $w(\cdot)$, for any vector u , there will be one and only one point E^u on the line of perfect equality, such that $w(E^u) = w(u)$. Moreover, given two points on the line of perfect equality, r and s , $w(r) \geq w(s)$ iff $\Sigma r \geq \Sigma s$. Therefore, for any two vectors u and v , $w(u) \geq w(v)$ iff $\Sigma E^u \geq \Sigma E^v$.

Further, because $w(\cdot)$ has a preference for perfect equality, $\Sigma E^u < \Sigma u$ if u is off the line of perfect equality. (To see this, consider z , such that $\Sigma z = \Sigma u$ and z is on the line of perfect equality. If u is not on the line of perfect equality, then $w(z) > w(u)$. Because $w(E^u) = w(u)$, it follows that $w(z) > w(E^u)$ and thus that $\Sigma z > \Sigma E^u$. But $\Sigma z = \Sigma u$, and thus $\Sigma u > \Sigma E^u$). Of course, if u is on the line of perfect equality, then $u = E^u$ and $\Sigma u = \Sigma E^u$.

Now we are ready to construct an inequality metric. Define $I^w(u)$ as $[1 - \Sigma E^u / \Sigma u]$. Note that this *is* a minimal inequality metric: it is greater than zero if u is off the line of perfect equality, otherwise zero. Further, it is clearly anonymous, because if u and v are permutations of each other, $E^u = E^v$.

It also can be seen that determining the total utility in a vector, and multiplying by its degree of equality, measured by $1 - I^w$, produces the very same ordering of vectors as $w(\cdot)$. In other words: $w(u) \geq w(v)$ iff $(\Sigma u)(1 - I^w(u)) \geq (\Sigma v)(1 - I^w(v))$. Why? Consider that $(\Sigma u)(1 - I^w(u)) = \Sigma u(1 - (1 - \Sigma E^u / \Sigma u)) = \Sigma E^u$. So $(\Sigma u)(1 - I^w(u)) \geq (\Sigma v)(1 - I^w(v))$ iff $E^u \geq E^v$. But we established earlier that, in turn, $E^u \geq E^v$ iff $w(u) \geq w(v)$.

x is at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $w(u(x)) \geq w(u(y))$

produces exactly the same ordering of outcomes as an SWF of the form:

x is at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} ,

$$\left(\sum_{i=1}^N u_i(x)\right)(1 - I^w(u(x))) \geq \left(\sum_{i=1}^N u_i(y)\right)(1 - I^w(u(y)))$$

where $I^w(\cdot)$ is a minimal inequality metric, if $w(\cdot)$ is continuous, Paretian, anonymous, and prefers a perfectly equal distribution of well-being.

This representation result *includes* continuous prioritarian SWFs, which set $w(\cdot)$ equal to $\sum_{i=1}^N g(u_i(x))$. In this case, the inequality metric corresponding to such SWFs is not just a minimal inequality metric, but an inequality metric in the full sense normally discussed by economists.⁷⁶ In particular, it includes Atkinsonian SWFs, which will yield the very same ranking of outcomes as overall well-being multiplied by the degree of equality, as measured via the Atkinsonian inequality metric.⁷⁷

In short, a very wide family of SWFs – *including* separable, Pigou-Dalton SWFs which satisfy a continuity requirement – can be *represented* as the upshot of amalgamating a concern for within-outcome fairness with overall well-being.⁷⁸

But this very interesting representation result does not resolve the question, which SWF is the within-outcome conception most plausibly understood to *justify*? It certainly does not

⁷⁶ In other words, in the case where $w(u(x)) = \sum_{i=1}^N g(u_i(x))$, $I^w(u(x))$ constructed as per the discussion in footnote 75 will satisfy the Pigou-Dalton principle as well as anonymity and equality-as-a-lower bound.

⁷⁷ The Atkinson inequality index is a mainstay of the literature on income inequality measurement. Applied to

utility vectors, it is defined as follows: $I^{Atkinson}(u(x)) = 1 - \frac{N}{\sum_{i=1}^N u_i(x)} \left(\frac{1}{N} \sum_{i=1}^N u_i(x)^{1-\gamma} \right)^{\frac{1}{1-\gamma}}$. The Atkinsonian

SWF, which this chapter ultimately endorses, says: x is morally at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} ,

$\frac{1}{1-\gamma} \sum_{i=1}^N u_i(x)^{1-\gamma} \geq \frac{1}{1-\gamma} \sum_{i=1}^N u_i(y)^{1-\gamma}$. But this yields the very same ordering of outcomes as the rule: x is

morally at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $\left[\sum_{i=1}^N u_i(x) \right] \left[1 - I^{Atkinson}(u(x)) \right] \geq$

$\left[\sum_{i=1}^N u_i(y) \right] \left[1 - I^{Atkinson}(u(y)) \right]$.

⁷⁸ Weirch (1983) uses the term “weighted utilitarianism” to refer to what is now known as prioritarianism, and argues that it arises by combining a concern for overall well-being and for equality.

mean that the within-outcome conception of fairness is most plausibly understood to justify a separable, Pigou-Dalton SWF.⁷⁹ To begin, why think that our evaluation of the distribution of well-being within outcomes should satisfy separability? For example, in the following sort of case, doesn't it seem intuitive that y' is not a more equal distribution than x' , even though y is a more equal distribution than x – in violation of separability?

Outcomes (listing individual utilities)

		\underline{x}	\underline{y}	\underline{x}'	\underline{y}'
<i>Individuals</i>	Audrey	93c	100c	93c	100c
	Baker	97c	100c	97c	100c
	Charlie	100c	100c	5c	5c
	Doris	100c	100c	5c	5c
	Ernie	100c	100c	5c	5c
	Flo	110c	100c	110c	100c

Further, there is no particular reason to think that a Pigou-Dalton transfer should always improve our assessment of within-outcome fairness. Here, it is important to distinguish between cases where the population size N is two, and larger populations. In the two person case, Pigou-Dalton transfers are clearly equality-improving. But why is that necessarily so with larger populations?⁸⁰ For example, consider the following case, in which a series of Pigou-Dalton transfers ends up with a single individual stranded at the bottom.

Outcomes (listing individual utilities)

		\underline{r}	\underline{s}	\underline{t}	\underline{w}	\underline{x}	\underline{y}
<i>Individuals</i>	Audrey	10c	10c	10c	10c	10c	10c
	Baker	20c	20c	30c	45c	45c	45c
	Charlie	30c	30c	30c	30c	45c	45c
	Doris	40c	40c	30c	30c	30c	45c
	Ernie	50c	60c	60c	45c	45c	45c
	Flo	60c	60c	60c	60c	45c	45c
	Gertie	70c	60c	60c	60c	60c	45c

In this example, each outcome is transformed into the outcome to its right via a Pigou-Dalton transfer. So the Pigou-Dalton principle means that outcome y is better than x , x better than w , w better than t , t better than s , and s better than r . Given transitivity, y is better than r . But it is hard to see why the distribution of well-being within y is really better than the distribution within r . The proponent of the within-outcome conception of fairness might resist this conclusion, without the drastic step of relinquishing transitivity, by repudiating the Pigou-Dalton principle.

⁷⁹ After all, if $w(\cdot)$ is a continuous, Paretian, anonymous function that prefers a perfectly equal distribution of utility but fails separability or Pigou-Dalton, the representation result will also apply to an SWF that uses $w(\cdot)$.

⁸⁰See Tungodden (2003; 20-21).

The proposition that the most plausible way to measure well-being inequality need not satisfy the Pigou-Dalton principle may be shocking to economists.⁸¹ Nor, I should say, is that proposition essential to my argument in this chapter (by contrast with the claim that the most plausible way to measure well-being inequality violates separability, which *is* essential to my argument). Still, the proposition is very interesting in its own right.

I should note that Temkin *concur*s in both these propositions. He agrees that the ranking of outcomes in terms of “comparative fairness” can fail separability and Pigou-Dalton.⁸² He argues to that effect, not just by appealing to intuitions, but by providing a systematic account of how to measure within-outcome fairness – an account that looks, within each outcome, to each individual’s “complaint” against individuals who are better off than her in that outcome. Why this can yield violations of separability and Pigou-Dalton is further discussed in the margin.⁸³

⁸¹ See the discussion of inequality metrics in Chapter 2.

⁸² See, for example, Temkin (2003b), which denies separability at various junctures (including, implicitly, in the “space traveler” example); and Temkin (1991; 83-84), questioning the Pigou-Dalton principle.

⁸³ For Temkin’s “complaint” view, see Temkin (1991, particularly chapters 2-3); Devooght (2003).

Temkin suggests (1) three possible views of who has a complaint, with corresponding baselines (that each person below the mean does; that each person except for the best-off person has a complaint relative to the best-off person; and that each person has a complaint against each person who is better off than him); and (2) three possible procedures for determining which outcome is worse in light of its complaints (an “additive” approach, which uses the shortfall between each individual’s well-being and the baseline as the measure of his complaint, and sums complaints; a “maximin” approach, which uses that shortfall as the measure of each individual’s complaint, but ranks outcomes depending on the largest complaint; and a “weighted additive approach,” which measures complaint using some measure which is non-linear in the well-being shortfall and then sums complaints, so as to give greater weight to individuals with larger shortfalls).

Without reviewing all the possibilities, it is easy to see that these various approaches can yield violations of separability and the Pigou-Dalton principle. Consider some illustrative examples. In each case vectors list the utility of different individuals in the population. All these examples involve cases in which the pairs of outcomes being compared have the same total well-being, so that the all-things-considered ranking of the outcomes reduces to their comparative fairness. *Relative to the Mean and Additive Procedure:* Let $x = (10c, 65c, 135c)$ and $y = (20c, 55c, 135c)$. Let $x^* = (10c, 65c, 15c)$, $y^* = (20c, 55c, 15c)$. Note that the Pigou-Dalton principle requires that y be better than x and that y^* be better than x^* . However, if we use this variant of the complaint approach, x gives individual 1 a complaint of $60c$ and individual 2 a complaint of $5c$, while y gives individual 1 a complaint of $50c$ and individual 2 a complaint of $15c$ (the means in both cases are $70c$). Because the sum of the complaints is $70c$ in both cases, this variant of the complaint procedure says both outcomes are equally good, violating Pigou-Dalton. Note also that it says that y^* is better than x^* , which violates separability. *Relative to the Mean and Weighted Additive Procedure.* In particular, let us measure each complaint as the square of the complainant’s shortfall, and then compare outcomes by summing these. To see how this can violate both separability and Pigou-Dalton, consider that a Pigou-Dalton transfer among individuals below the mean will be an improvement, but that a Pigou-Dalton transfer above the mean will not. For example, if $x = (400c, 80c, 120c)$, a Pigou-Dalton transfer yields $y = (400c, 90c, 110c)$. Summing squared shortfalls from the mean assigns a score of $14,400c + 6400c = 20,800c$ to x , and a score of $12,100c + 8100c = 20,200c$ to y , so y is ranked better. On the other hand, if $x = (10c, 80c, 120c)$ and $y = (10c, 90c, 110c)$, the mean is $70c$ and so the procedure now on the table ranks the outcomes as equally good, in violation of both Pigou-Dalton and separability. *Relative to the Best Off Person and Additive Procedure.* Let $x = (10c, 50c, 100c)$. A Pigou-Dalton transfer yields $y = (10c, 60c, 90c)$. Summing complaints relative to the best-off individual assigns a score of $90c + 50c$ to x , and a score of $80c + 30c$ to y . So y is ranked better. On the other hand, if $x^* = (200c, 50c, 100c)$ and $y^* = (200c, 60c, 90c)$, x^* is assigned a score of $150c + 100c$, and y^* is assigned a score of $140c + 110c$, so the two are ranked equally good, in violation of both Pigou-Dalton and separability. *Relative to the Best off and Weighted Additive Procedure.* Let $x = (10c, 100c, 100c, 120c)$ and $y = (10c, 90c, 115c, 115c)$. Then summing

Which Conception is More Plausible?

Nothing in the analysis to this point provides an affirmative case in favor of the across-outcome conception of fairness or in favor of prioritarian (Pigou-Dalton, separable) SWFs. I have argued that certain conjunctions of propositions “hang” together, and that others do not. An across-outcome conception of fairness coheres well with a prioritarian SWF (but not an SWF that fails either separability or Pigou-Dalton); a within-outcome conception coheres well with an SWF that fails both Pigou-Dalton and separability (but not an SWF that possesses either characteristic). This suggests that *either* the conjunction of an across-outcome conception of fairness with a prioritarian SWF, *or* the conjunction of a within-outcome conception of fairness with a non-separable and non-Pigou Dalton SWF, is a plausible initial candidate for a point of reflective equilibrium. How shall we adjudicate *between* these two positions?

By appealing to welfarism. Welfarism, itself, better coheres with the across-outcome conception of fairness than with the within-outcome conception. Welfarism, remember, is a person centered moral view; it is consequentialist, in rejecting agent-relative constraints and prerogatives and being oriented around an impartial ranking of outcomes; and it focuses on human well-being, rather than other human attributes that might be thought to have moral relevance. Its person-centeredness and well-being focus are both formally captured in axioms of Pareto superiority and indifference. I suggest that the across-outcome conception of fairness is one part of a simple and tightly unified set of propositions that explain why morality is welfarist.

(1) *Person-centeredness and fairness*: The focus of morality is what persons owe to each other.⁸⁴ Other beings may be a source of non-moral reasons. Decisionmakers may have non-moral reason to protect animal well-being, protect the environment, promote beauty, and so forth. But persons, uniquely, are capable of being guided by normative deliberation. This yields a distinct subset of normative principles – *moral* norms – which are the norms that bind each member of a community of persons, in virtue of reflecting a fair and impartial concern for all of their interests.

(2) *Consequentialism*: Moral decisionmaking should be structured around an impartial

squared shortfalls relative to the best off says that y is better than x . But let $x^* = (200c, 100c, 100c, 120c)$ and $y = (200c, 90c, 115c, 115c)$. Summing squared shortfalls says that x^* is better than y^* . This violates separability. *Maximin*. Imagine that one individual, Sue, has the largest complaint (whether measured by her well-being shortfall from the baseline or some non-linear function; and whether each individual has a complaint relative to the mean, the best off individual, or all those better off than her). It is not hard to see that a Pigou-Dalton transfer between two other individuals, from Fred to Jim, might leave Sue's complaint unchanged (so that the maximin complaint rule ranks the outcomes as equally good, in violation of Pigou-Dalton); and that changing Sue's well-being level in both outcomes so that it is now Jim who has the largest complaint in both outcomes, which is diminished by the Pigou-Dalton transfer -- so that there is a violation of separability. (Admittedly, a maximin rule may well be less plausible than a leximin rule for comparing arrays of complaints, and both less so than additive or weighted additive approaches).

⁸⁴ This idea is characteristic of the Kantian tradition in moral philosophy, represented by various contemporary philosophers, such as Rawls, Nagel and, most recently, Scanlon.

ranking of outcomes. An impartial ranking of outcomes must be “agent neutral” rather than “agent relative” (it must not depend on the identity of the decisionmaker) and must satisfy additional constraints, such as anonymity. Morality is oriented around outcomes because all rational choice has this feature (a fundamental idea captured by EU theory). The moral ranking of outcomes is *impartial* because morality is the domain of normativity characterized by an impartial concern for humans’ interests; and in particular because norms of fairness should be impartial as between all members of the community meant to be guided by those norms.

(3) *Across-Outcome Conception of Fairness*: As above. The moral ranking of any pair of outcomes should be determined by balancing individuals’ claims in favor of one or another, consistent with the basic claim/well-being valencing rules. Well-being is the appropriate “currency” for claims, because an individual’s well-being is just the same as the “goodness of outcomes’ *for* him, or the goodness of outcomes “from his perspective.” Modulo considerations of responsibility, it is fair that an individual have a claim in favor of an outcome iff he is better off in that outcome, and that he have no claim either way iff he is equally well off.

(4) *Pareto Indifference and Superiority*. The moral ranking of outcomes should satisfy Pareto indifference and superiority. This follows from (3).

By contrast, the within-outcome conception does not cohere as well with welfarism. In particular, it does not cohere as well with the principle of Pareto superiority. Imagine that one sees the inequality of well-being within outcomes as helping to justify the moral ranking of outcomes. As already explained, within-outcome fairness, alone, cannot yield a ranking of outcomes that satisfies Pareto superiority: a Pareto-inferior move from an unequal distribution to a perfectly equal distribution will decrease the degree of inequality and thus increase fairness in the within-outcome sense. Thus, in order to explain why the moral ranking of outcomes satisfies Pareto superiority, the proponent of the within-outcome conception of fairness must combine fairness (in that sense) with some other consideration – in particular, overall well-being. But why believe that the result of balancing within-outcome fairness and overall well-being will *necessarily* satisfy the Pareto principle?

The following example illustrates the issues, here. Imagine that there is some person in the population, call him Trump, who is much better off than everyone else in the world. You are a space traveler, zooming by the world, and in your travels you happen to have picked up an object that has sentimental value for Trump, but is useless to anyone else (including you).⁸⁵ For Trump to possess the object would make his life even better. Pareto superiority requires that you give Trump the object. From the perspective of an across-outcome view, it is clear why this is so. Trump has a genuine interest in the object; everyone else is unaffected; whatever the moral

⁸⁵ This example builds on Temkin’s space traveler case. See Temkin (2003b; 68-69).

rules for resolving conflicting claims, this is a no-conflict case, in which all non-zero claims point in the same direction. By contrast, from the perspective of a within-outcome view, it is far from clear why you should give Trump the object. Making him better off (presumably) makes an inequitable distribution of well-being even worse. Why be sure that the increase in Trump's well-being necessarily trumps the worsening in the inequality of well-being produced by giving him the object?

To be clear, I am *not* suggesting that the proponent of the within-outcome conception is logically compelled to violate Pareto superiority in this case or any other. Not at all. The proponent of within-outcome fairness *can* formulate a moral view that combines fairness (thus conceived) with overall well-being, and that is not only person-centered, consequentialist and satisfies Pareto indifference, but is also welfarist by virtue of including Pareto superiority as an independent, underived axiom. The point, rather, is that such a view is less simple and coherent than one which sees Pareto indifference and superiority as flowing directly from the idea of fairly reconciling individuals' claims. Such a view is forced to construe person-centered morality as a hybrid of overall well-being and fairness, even though the strongest case for making morality person-centered flows from the unique status of persons vis a vis the concept of fairness⁸⁶; and it is forced to posit, without a deeper explanation, that the balancing of fairness and overall well-being must always yield a ranking of outcomes that satisfies Pareto superiority.

My analysis, here, hardly provides a slam dunk case against the variant of welfarism that incorporates a within-outcome conception of fairness, and thus justifies an SWF that fails separability and Pigou-Dalton. There is no logical inconsistency in such a view. It is in the nature of the coherentist reasoning that characterizes much of moral philosophy to reach contestable conclusions, perhaps unsatisfyingly so. Still, I think a good case can be made that the across-outcome conception of fairness fits better with welfarism⁸⁷ – and thus in turn that the most attractive way to flesh out welfarism, using the SWF framework, is to use the sort of SWF that this conception justifies, namely a Pigou-Dalton and separable one.

This conclusion, I should note, is connected to an important strand in the literature on prioritarianism, which focuses on what Temkin calls the “person affecting” principle (Temkin calls it “the slogan”) and on the so-called “leveling down” objection. I discuss this connection in the margin.⁸⁸

⁸⁶Again, by contrast, it is not true that only persons can possess a well-being, or that the idea of overall well-being must be limited to persons – although it is true that persons' well-being is qualitatively distinct from animals', so that the idea of a person-centered moral view which is a hybrid of the overall well-being of persons and fair distribution is certainly not wholly incoherent.

⁸⁷Indeed, Temkin might well accept this conclusion. He rejects welfarism and rejects or at least calls into question the Pareto principle. See Temkin (1991; 139-40; 2003b).

⁸⁸On these issues, see Parfit (1991); the literature on prioritarianism cited supra note __; Temkin's scholarship cited supra note __; and also __.

The person-affecting principle has many different variants. Roughly, it says something like the following: outcome x is not better than y (either in the all-things considered sense, or in even one respect) unless it is better for

Assume that I have made the case that welfarism better coheres with the across-outcome conception of fairness than with the within-outcome conception. The reader might see this observation as yielding an additional argument *against* welfarism. The standard case against welfarism takes issue with its *consequentialism* (pointing to the putative existence of agent-relative constraints and options), and with its focus on *human well-being* (arguing that non-human beings and entities have moral relevance, or that the currency for assessing distribution is something other than well-being, e.g., resources). The non-welfarist reader might now advance an additional argument, along the following lines. Because the principle of Pareto superiority sits uncomfortably with our natural tendency to think about fairness in within-outcome terms – our natural tendency to focus on the inequality of well-being or other currencies within outcomes – we should reject that principle, thus reject welfarism, and stick with the within-outcome conception of fairness.

Chapter 1 briefly reviewed the standard case against welfarism, trying to show how the standard objections might be countered. However, as explained in that chapter, the main project of this book is to work within welfarism – to determine which framework for systematically evaluating government policies and other large-scale choices is the most attractive specification of welfarism. That enterprise should be of interest, of course, for welfarists, but it should also be of interest for nonwelfarists (for reasons reviewed in Chapter 1).

What of this new objection to welfarism? I believe that the objection fails. To my mind, the principle of Pareto superiority is intuitively compelling, even in cases like the Trump case. How can making one individual worse off produce a morally better outcome? And thinking about fairness as a matter of balancing the claims of affected persons, ignoring the unaffected, strikes me as no less natural than the within-outcome approach.

some person (again, either in the all-things-considered sense, or in at least one respect). Scholarly discussion of different specifications of the principle, and their plausibility, has grown quite complicated, and I cannot review it here. In addition to the general literature on prioritarianism, *supra* note __; and Temkin's scholarship, *supra* note __, see Brown (2003); Christiano & Braynen (2008); Doran (2001); Holtug (2007b); Mason (2001); O'Neill (2008); Ramsay (2005).

Note, however, that the across-outcome conception of fairness, as I have presented it, is a *person-affecting* conception, in the following sense: An individual has a claim in favor of outcome x over y just in case the obtaining of x rather than y has an all-things-considered positive effect on her, i.e., increases her well-being. Fairness (on this view) is a matter of adjudicating between these claims, with the minimal requirement that one outcome is fairer than another only if there is at least one person who has a claim in its favor, i.e., would be all-things-considered positively affected by the obtaining of that outcome, or at least an incomparable claim.

The leveling-down objection to egalitarianism, as discussed in the literature, basically runs as follows: in the case where x is Pareto inferior to y , and well-being is more equal in x , how is it possible that x is better than y in any respect, let alone all-things-considered? The argument that I present in the text for the across-outcome view is connected to the leveling-down objection, in the following way: The Pareto principle itself does not rule out the possibility that x is better than y in one respect -- because well-being is more equal -- even though x is Pareto inferior. But what is difficult to explain is why – if we see a concern for well-being equality as one component of the ranking of outcomes – we would not also be prepared to conclude in some cases that a Pareto-inferior move is all-things-considered better, which *does* of course violate the Pareto principle.

Once more, however, it is simply too ambitious for me to mount a real defense of welfarism in this book. I cannot hope, in a single book, both to show how welfarism is best specified, and to fully rebut objections *to* welfarism, standard or new. I am therefore content to have reached the following, conditional, conclusion in this section. *If* welfarism is the most attractive moral view, and if welfarism should be structured using the SWF framework, *then* the most attractive moral decision procedure employs a prioritarian SWF. An ancillary result of my analysis is that inequality metrics have no useful role to play in welfarist decisionmaking.⁸⁹

Sufficientism

Roger Crisp’s recent work on “sufficientism,” in his influential article “Equality, Priority, and Compassion,” points to the possibility of an SWF that is separable but fails the Pigou-Dalton principle.⁹⁰ I will argue that such an SWF is unattractive. The sufficientist SWF is not the SWF which is best justified in light of *either* understandings of fairness mooted in this chapter – not in light of the within-outcome conception, because it is separable; and not in light of the across-outcome conception, because it fails Pigou-Dalton.

In his article, Crisp criticizes prioritarianism on two counts. One criticism has to do with aggregation. Crisp suggests that standard prioritarian schemes for adjudicating between better and worse-off individuals either give too little, or too much, weight to better-off individuals. Crisp criticizes an “absolute priority” approach (exemplified by a leximin SWF⁹¹), because it prefers giving a small benefit to one badly off individual, to giving large benefits to many individuals who are only slightly better off. On the other hand, Crisp rejects a “weighted priority” view (exemplified by a continuous prioritarian SWF, which uses the formula

$\sum_{i=1}^N g(u_i(x))$), because it would forego a substantial benefit to a badly off individual, in order to produce a very small gain to a sufficiently large group of well-off persons. For example, he

⁸⁹ Given the large literature on inequality metrics, this conclusion is certainly important in its own right. Why reach this conclusion? The most attractive welfarist procedure uses an SWF – or so I generally argue in this book. It is certainly true that -- as we saw earlier-- many distribution-sensitive SWFs *can* be expressed as the product of a measure of equality and overall well-being, including prioritarian SWFs. But if the across-outcome conception of fairness fits best with welfarism, and the SWF is in turn best justified by that view of fairness, this decomposition of the SWF is misleading. According to the across-outcome conception, the degree of well-being inequality within outcomes is *not* what determines the fairness of outcomes; and the degree of well-being inequality, together with overall well-being, is *not* what justifies the conclusion that one outcome is morally better than another.

⁹⁰ Crisp (2003). See also Crisp (2006; 146-62); Frankfurt (1987).

⁹¹ Strictly, Crisp criticizes giving absolute priority to the worst-off person. That is achieved, not just by the leximin SWF, but also others. However, leximin is the most straightforward example of a Paretian SWF that gives absolute priority to the worst off. (Note that maximin, another standard social choice rule which gives absolute priority to the worst off, is not Paretian.) Moreover, if Crisp finds giving absolute priority to the worst off problematic, then *a fortiori* he would find leximin problematic in virtue of giving absolute priority to worse off persons, even when they are not worst off. On these issues, see Brown (2005, chapter 5); Brown (2005b).

observes, the “weighted priority” approach might forego pain relief for 10 very badly off individuals (which would move each of them from welfare level 1 to welfare level 50), so as to give a nice chocolate candy to 15,000 rich individuals (which would move each of them from level 98 to level 99).

Crisp’s second criticism is that giving priority to the well-being of the worse off is implausible when all the individuals involved are sufficiently well off. Crisp illustrates this point through “the Beverly Hills case,” in which the prioritarian prefers to give fine wine to a small group of rich individuals as opposed to giving fine wine to a larger group of super-rich individuals, where the individual welfare benefit of fine wine is the same amount for all the individuals in both groups.

It seems somewhat absurd to think that the *Rich* should be given priority over the *Super-Rich* [W]hat the Beverly Hills case brings out is that, once recipients are at a certain level, any prioritarian concern for them disappears entirely. This implies that any version of the priority view must fail: when people reach a certain level, even if they are worse off than others, benefitting them does not, in itself, matter more. [E]ven if the benefits to each of the *Rich* and the *Super-rich* are identical and their numbers are the same, there still seems to be nothing to be said for giving priority to the ‘worse off.’ At this level, only utilities matter ...⁹²

These two criticisms lead Crisp to propose a moral theory that incorporates a well-being threshold, identified as the level such that an impartial spectator would feel compassion for individuals below but not above it. He endorses “The Compassion Principle”:

The Compassion Principle: Absolute priority is to be given to benefits to those below the threshold at which compassion enters. Below the threshold, benefitting people matters more the worse off those people are, the more of those people there are, and the greater the size of the benefit in question. Above the threshold, or in cases concerning only trivial benefits below the threshold, no priority is to be given.⁹³

Crisp entertains, but rejects, the suggestion that the compassion threshold is at the level where individuals’ basic needs are satisfied.

A problem with the [need] proposal is that, on any plausible distinction between needs and, say, desire satisfaction or other components of welfare, needs give out before compassion. Imagine a society which includes, among a large number of very wealthy and flourishing individuals, a group which is very poor but whose basic and indeed nonbasic needs are met. Compassionate concern for the badly off speaks in favor of at least some transfers from the rich to the poor, even if the poor use any resources gained to purchase goods which they could not be said to need.⁹⁴

Instead, the compassion level is to be understood as the level of a life which is “sufficiently good,” which Crisp specifies as 80 years of high-quality life on earth.

⁹² Crisp (2003; 755).

⁹³ *Id.* at 758.

⁹⁴ *Id.* at 759.

Imagine that the impartial spectator knows that the universe contains trillions of beings whose lives are at a much higher level of welfare than even the best off on this planet. Will he or she take the same view of the Beverly Hills case, or might his or her threshold for compassion be set at a much higher level? It is hard to know how to answer such questions, but, on reflection, my own intuition is that, say, eighty years of high-quality life on this planet is enough, and plausibly more than enough, for any being.⁹⁵

Campbell Brown has cogently formalized Crisp’s *Compassion Principle* as a type of SWF.⁹⁶ For short, let us term this the “sufficientist” SWF. Crisp formulates this SWF for the case in which there is a single utility function that tracks well-being; I will generalize his idea to the case in which there is a set \mathbf{U} of utility functions that represents the well-being associated with life histories.

For a given outcome set, identify a life-history, $(x^*; i^*)$ that lies at the “compassion” threshold. For each $u(\cdot)$ belonging to \mathbf{U} , set a utility threshold (for that function) equal to the utility of the threshold life-history. Now compare each pair of outcomes, x and y , using the following two-stage approach. Take some strictly increasing, strictly concave $g(\cdot)$ function. For each utility function $u(\cdot)$ in \mathbf{U} , each outcome corresponds to a below-threshold utility vector, namely the vector of individual utilities truncated at the level of the utility threshold. Each outcome also corresponds to an above-threshold utility vector, namely the vector of individual utilities or the threshold utility, whichever is larger. For each utility function, assign outcomes a primary score by summing the elements of the below-threshold vector, transformed by the $g(\cdot)$ function. Then assign outcomes a secondary score by summing the elements of the above-threshold vector, without any weights. If one outcome has a higher primary score it is better according to $u(\cdot)$; if two outcomes have the same primary score and one has a higher secondary score, it is better according to $u(\cdot)$; otherwise the two outcomes are equally good according to $u(\cdot)$. Finally: an outcome is at least as good as outcome y , full stop, iff it is at least as good for all $u(\cdot)$ belonging to \mathbf{U} .⁹⁷

This “sufficientist” SWF has exactly the features Crisp argues for. It conforms to the aggregation approach he recommends: it gives absolute priority to below-threshold individuals as against above-threshold individuals (unlike the continuous prioritarian SWF), but does make tradeoffs in giving benefits to different groups all of whom are below the threshold (unlike leximin). Further, as Crisp recommends in his discussion of the “Beverly Hills” case, it

⁹⁵ *Id.* at 762.

⁹⁶ See Brown (2005, chapter 5; 2005b).

⁹⁷ Formally, for each $u(\cdot)$, the threshold value $U^* = u(x^*; i^*)$, where $(x^*; i^*)$ is the threshold life history. Define an N entry vector $u^+(x)$ as follows: Its i th entry is $u(x; i)$ if $u(x; i) \leq U^*$, otherwise the entry is U^* . Define an N entry vector $u^{++}(x)$ as follows: Its i th entry is $u(x; i)$ if $u(x; i) \geq U^*$, otherwise the entry is U^* . Outcome x is better than y

according to $u(\cdot)$ iff: (1) $\sum_{i=1}^N g(u_i^+(x)) > \sum_{i=1}^N g(u_i^+(y))$ or (2) $\sum_{i=1}^N g(u_i^+(x)) = \sum_{i=1}^N g(u_i^+(y))$ and

$\sum_{i=1}^N u_i^{++}(x) > \sum_{i=1}^N u_i^{++}(y)$. Outcomes x and y are equally good according to $u(\cdot)$ if neither is better than the other.

eschews giving priority to benefits accruing to worse off individuals, as against benefits accruing to better off individuals, when all the individuals involved are better off than the threshold life history.

The sufficientist SWF, it should be noted, is separable. However, it fails the Pigou-Dalton principle. This violation of the Pigou-Dalton principle, of course, is a logical implication of the fact that the sufficientist SWF does not prioritize benefits to worse-off individuals who are above the well-being threshold. If it *did* satisfy the Pigou-Dalton principle, then it would necessarily give priority to benefits accruing to worse-off individuals over benefits accruing to better-off individuals, regardless of how well off all the individuals might be.

Is the sufficientist SWF attractive?⁹⁸ I argued earlier that the within-outcome conception of fairness most plausibly justifies a non-separable SWF. So it is hard to see how fairness, thus understood, would justify the sufficientist SWF. Indeed, Crisp vigorously criticizes the within-outcome conception; and Temkin, the chief contemporary proponent of that approach, has in turn criticized sufficientism.⁹⁹

What about an across-outcome conception? If outcomes are ranked as more or less fair by seeing individuals as having claims for or against outcomes, might that warrant a sufficientist SWF?

The key difficulty, as I see it, is in explaining why an across-outcome approach would justify a well-being threshold above which the Pigou-Dalton principle fails. One possibility is to say that the threshold is the ceiling for individuals' claims. Each individual has a claim to be made better off, up to the threshold.¹⁰⁰ This approach would mean rejecting the basic connection between an individual's well-being and claim valence which I have "built into" my statement of the across-outcome conception. It would mean this: If, for a given pair of outcomes, there are some affected individuals who are above a threshold in both outcomes, then those individuals have no claim either way. At the extreme: If, for a given pair of outcomes x and y , all affected individuals are above the threshold in both outcomes, then no individuals have a claim either way.¹⁰¹ This approach is quite implausible. If I would be genuinely better off in x than y , then

⁹⁸ For discussion of sufficientism, see the sources cited elsewhere in this section and also Benbaji (2005); Casal (2007).

⁹⁹ Crisp (2003; 745-50; 2006, 146-52); Temkin (2003a).

¹⁰⁰ To formalize this interpretation: Given two outcomes x and y and a utility function $u(\cdot)$, each individual's claim in favor of x is measured by a single number, namely $g(u^+(x; i)) - g(u^+(y; i))$, where $u^+(\cdot; i)$ is the i th element of the truncated u^+ vector, as defined in footnote 97. On this interpretation, outcome x is fairer than y according to $u(\cdot)$ if the sum of claims is positive; equally fair if this sum is zero; and less fair than y if this sum is negative. And one outcome is at least as fair as another iff it is at least as fair, for all $u(\cdot)$ belonging to \mathbf{U} .

¹⁰¹ This approach does not see an increase in the well-being of a person already above the threshold as making the outcome more fair. So, for the proponent of this approach to avoid a violation of Pareto superiority, she would need to see morality as the combination of fairness and overall well-being (and we can ask her, as with the discussion earlier of the comparative fairness approach, why overall well-being is a *moral* consideration.). Note also that the combination of fairness in this sense and overall well-being does not justify the sufficientist SWF as I have defined

(modulo considerations of responsibility), doesn't I have a moral interest in x rather than y obtaining?

A different possibility is that the connection between the well-being and claim valence is preserved, but that the threshold is used in calibrating the strength of individuals' claims. See the margin for a full statement of this approach, and how it would be generate the sufficientist SWF.¹⁰² As can there be seen, this approach has the implication that, if an individual is above the threshold in both outcomes x and y , she *does* have a claim in favor of the outcome in which she is better off; but the measure of the strength of that claim is simply her well-being difference between the outcomes.¹⁰³ Thus, if we are comparing a pair of outcomes x and y , and all affected individuals are above the threshold, then x is fairer than y just in case the sum of well-being differences of the affected individuals is positive. And this means, in particular, that if i and j are above the threshold in both x and y , and everyone else is unaffected, and the difference in i 's well-being between x and y is the same as the difference in j 's well-being between y and x , and i is worse off in both outcomes, then the two individuals have equally strong claims and the outcomes are equally fair.

But why countenance this violation of the Pigou-Dalton principle? Whatever your theory of well-being, construct in your mind a case in which the two individuals, i and j , are very well-off in both outcomes, with everyone else unaffected. However, neither individual is at the very top of the ladder of human well-being (however you construct that ladder). Individual i genuinely stands to benefit from outcome x rather than y obtaining, and individual j genuinely stands to benefit the very same amount from y rather than x obtaining. Moreover, j is higher on the ladder of genuine well-being (as you see it) in both outcomes. Modulo considerations of responsibility, isn't it *fairer* that outcome x rather than outcome y obtain?

Crisp's Beverly Hills case is not a particularly helpful case for thinking clearly about the answer to this question. In considering the possibility of thresholds, for purposes of welfarism, it

it, but instead leads to a slightly different SWF which applies the utilitarian formula to $u(x)$ and $u(y)$ rather than $u^{++}(x)$ and $u^{++}(y)$.

¹⁰² On this interpretation, for a given pair of outcomes and a utility function $u(\cdot)$, each individual i 's claim in favor of x is measured by a pair of numbers, $(g(u^+(x; i)) - g(u^+(y; i)), u^{++}(x; i) - u^{++}(y; i))$, where $u^+(\underline{\cdot}; i)$ is the i th element of the u^+ vector and $u^{++}(\underline{\cdot}; i)$ is the i th element of the u^{++} vector, as defined in footnote _____. Call the first number the primary measure, the secondary measure. Then outcome x is fairer than y according to $u(\cdot)$ iff: (1) the sum of individuals' primary measures is positive; or (2) the sum of individuals' primary measures is zero and the sum of individuals' secondary measures is positive. Outcome y is fairer than x according to $u(\cdot)$ iff: (1) the sum of individuals' primary measures is negative; or (2) the sum of individuals' primary measures is zero and the sum of individuals' secondary measures is negative. The outcomes are equally fair according to $u(\cdot)$ if neither is the case. One outcome is at least as fair, full stop, iff at least as fair for all $u(\cdot)$ belonging to \mathbf{U} . It is clear that this interpretation of fairness generates the sufficientist SWF. The sufficientist SWF defines a rule for when x is better than y according to $u(\cdot)$ (see footnote 97 above), and that is true just in case the outcome x is fairer than y according to $u(\cdot)$, as just defined. Ditto for y being better than x , or the two outcomes being equally good.

¹⁰³ In the case of such an individual, the primary measure of her claim is zero, for all $u(\cdot)$, and her secondary measure is just her utility difference.

is vitally important not to conflate (1) a threshold *within* the most attractive account of well-being, which would show up as a threshold within the utility functions that map the various determinants of well-being (the various constituents of outcomes) onto utility numbers representing well-being, with (2) a threshold in the moral weight of well-being itself. But the *Beverly Hills* case invites just such a conflation. We are told that each Rich individual stands to benefit a certain amount by having fine wine, and that each Superrich individual stands to benefit the very same amount by having fine wine. But the thought experiment of giving fine wine to someone living in Beverly Hills, the lap of luxury, naturally invites the reaction that this individual is above a threshold of the first sort – in particular, that giving her this additional luxury good doesn't make her life genuinely better at all. As Nagel cogently observes (anticipating, and rejecting, a sufficientist approach years before Crisp's article): "My moral instincts reveal no egalitarian priority for the well-to-do over the rich and super-rich. But I suspect that is because the marginal utility of wealth diminishes so steeply in those regions ... that these categories do not correspond to significant objective differences in well-being"¹⁰⁴

Crisp might respond that he is offering a conception of *compassion*, not fairness. Our compassionate concern for people gives out at some threshold level of well-being. But – regardless of whether this is true, and regardless of whether morality should incorporate considerations of "compassion" – morality certainly should be sensitive to considerations of interpersonal *fairness*; the most attractive such conception, consistent with welfarism, is the across-outcome conception; and *that* conception argues in favor of the Pigou-Dalton principle, full stop, rather than allowing it to be violated above some threshold of well-being.¹⁰⁵

What about Crisp's views about aggregation: his criticism of leximin for giving absolute priority to badly off individuals, as against individuals who are only slightly better off; and his criticism of continuous prioritarian SWFs for allowing very small gains to sufficiently large numbers of well-off individuals to override a substantial benefit for a badly off person? As work by Campbell Brown and Bertil Tungodden has helped show, it is possible to respond to these criticisms *within the context of prioritarianism*.¹⁰⁶ It is possible to craft a SWF that (1) satisfies Pigou-Dalton and separability, hence is "prioritarian" as I am using that term, rather than being the sufficientist SWF; but (2) differs from both leximin and the continuous prioritarian SWF in its approach to aggregation, because it uses a threshold to mark the point of absolute priority. For short, I will term this a prioritarian SWF with an absolute threshold.

Nothing I have said, thus far, would rule out such an approach. Up to this point in the chapter, my argumentation has been focused on showing that the welfarist crafts a more coherent and unified moral view by conceiving of fairness in the across- rather than within-outcome

¹⁰⁴ (Nagel 1991, 70).

¹⁰⁵ See also Chapter 5, discussing whether intuitions concerning short-term human suffering and hardship that are inconsistent with a continuous prioritarian SWF applied to lifetime utilities should be credited and, if so, whether they should be seen as reflecting moral reasons.

¹⁰⁶ See below.

sense; and that the across-outcome conception of fairness, in turn, is most plausibly specified to require that the ordering of outcomes be separable and satisfy the Pigou-Dalton axiom. This analysis, if cogent, provides a case for prioritarian SWFs as a general family – as opposed to non-separable SWFs; or an SWF which is separable but fails Pigou-Dalton, such as the sufficientist SWF. Let us now see what can be said about the choice *within* the family of prioritarian SWFs.

Prioritarian SWFs

My ultimate aim in this chapter is to argue for a continuous prioritarian SWF – in particular the Atkinsonian SWF. The Atkinsonian SWF is one of the most widely used by economists who follow the SWF approach. But what argues for it? In this section, I will focus on making the case for a continuous prioritarian SWF. How we get from there to the Atkinsonian SWF will be covered below.

Again, what I mean by a continuous prioritarian SWF is one that says: x is at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $\sum_{i=1}^N g(u_i(x)) \geq \sum_{i=1}^N g(u_i(y))$, with $g(\cdot)$ strictly increasing and strictly concave. This SWF takes an additive approach to specifying the across-outcome conception of fairness. For each pair of outcomes, x and y , and each utility function $u(\cdot)$ in \mathbf{U} , each individual's claim in favor of x or y is measured by a single number; the fairer outcome according to $u(\cdot)$ is determined by summing these numbers; and the fairer outcome, full stop, is determined by seeing which is fairer across all utility functions in \mathbf{U} .¹⁰⁷

Because the continuous prioritarian SWF measures claims in this manner and adjudicates between conflicting claims by summing their measures, it has an aggregation property which many might find unattractive.¹⁰⁸ There are a variety of ways of precisely characterizing the sense in which this SWF is “aggregative.” I will focus on the following property (call it the “Numbers Win” property), which is meant to show this SWF in the least favorable light. The Numbers Win property is this:

Numbers Win: In any outcome set, consider any life-history $(x; i)$, however bad it may be, and any other life-history $(y; i)$, which is even worse for that individual i . Choose any number F larger than 1, which can be arbitrarily large; and any positive number b ,

¹⁰⁷ Formally: given a pair of outcomes x and y and a utility function $u(\cdot)$, the strength of each individual's claim in favor of x , according to $u(\cdot)$, is measured by $g(u(x; i)) - g(u(y; i))$, with a negative number meaning that i has a claim for y . Outcome x is at least as fair as y , according to $u(\cdot)$, iff this sum is nonnegative. Outcome x is at least as fair as y , full stop, if at least as fair according to all $u(\cdot)$ belonging to \mathbf{U} . An individual has a claim in favor of x iff his claim according to $u(\cdot)$ is nonnegative for all $u(\cdot)$ belonging to \mathbf{U} and positive for some; a claim in favor of y iff his claim according to $u(\cdot)$ is nonpositive for all $u(\cdot)$ belonging to \mathbf{U} and negative for some; and no claim either way iff the measure of his claim is zero for all $u(\cdot)$.

¹⁰⁸ More precisely, it has this property at least in the case where \mathbf{U} is formed by pooling a finite number of spectators' complete preferences over life histories, and the Atkinsonian SWF is used. See footnote 110 below.

which can be arbitrarily close to zero. Then for an SWF to have the “Numbers Win” property means: There is some sufficiently large positive number M , such that *if* the only individuals affected by y and x are (1) individual i , who goes from however badly off he might be in x , to being even worse off in y (by however much), and (2) at least M individuals each of whom in x is F times better off than individual i in x , and each of whom is better off in y than x , but only by an amount which is a fraction b of individual i 's loss, *then* the SWF counts y as a morally better outcome than x .

In other words, a large loss to a badly off individual might be overridable by arbitrarily small benefits to arbitrarily well-off individuals, no matter how small their benefits or well-off those other individuals may be, and no matter how badly off the first individual might be or substantial his loss.¹⁰⁹ Nothing in the structure of the continuous prioritarian SWF precludes such aggregation.¹¹⁰

¹⁰⁹ It should be assumed, however, that individual i in outcome x is better off than nonexistence. (Otherwise, individuals who are a positive number F times the utility of individual i will be worse off than him, not better off.) I could reformulate the property in a more complicated fashion to allow for a negative starting point, but will not belabor the discussion by doing so.

On this issue, note also that the most attractive continuous prioritarian SWF, the Atkinsonian SWF, is not applicable to negative utilities (see below) and thus can only be used to rank outcome sets in which each individual is at least as well off as nonexistence in all outcomes. Thus the Atkinsonian SWF has the Numbers Win property only in the sense that it is willing to make an arbitrarily badly off individual even worse off (but better off than nonexistence) for the sake of arbitrarily small benefits to many much better off individuals. Still, having the Numbers Win property even in this somewhat cabined sense still seems quite problematic.

¹¹⁰ I have formulated Numbers Win in terms of ratios (with well-off individuals an arbitrary multiple of individual i 's well-being in x , and their benefit an arbitrary small fraction of i 's loss) because, with utility at most measurable up to a positive ratio transformation, it is not clear what it means to identify a particular numerical level of well-being or size for a benefit.

I assume that $(x; i)$ is better than nonexistence. To say that some individual j is at an arbitrarily large multiple F of that life-history is to say that, for all $u(\cdot)$ belonging to \mathbf{U} , $u(x; j)/u(x; i) = F$. To say that individual j 's benefit from moving to outcome y is an arbitrarily small fraction b of i 's loss is to say that, for all $u(\cdot)$ belonging to \mathbf{U} , $(u(y; j) - u(x; j))/(u(x; i) - u(y; i)) = b$.

Imagine that \mathbf{U} is a singleton, which consists of a single utility function. In that case, for any continuous prioritarian SWF, the badly off individual has a claim to x , measured by $g(u(x; i)) - g(u(y; i))$. Call this c_i . Let us designate the magnitude of individual i 's well-being loss, if y were to obtain, i.e., $u(x; i) - u(y; i)$, as Δu . Imagine that x includes a group of individuals, all of whom are F times better off than individual i , and all of whom benefit from y by amount $b\Delta u$. Then each of those individuals has a claim in favor of y equaling $g(Fu(x; i) + b\Delta u) - g(Fu(x; i))$. Call this c^* . Because $g(\cdot)$ is an increasing function, this number is positive. Let M equal the first integer greater than c_i/c^* . If the group has M or more members, and the only individuals affected by the two outcomes are those individuals and individual i , then the SWF says that y is, on balance, a fairer outcome.

Note that, for a given pair of outcomes x and y , with i worse off in y than x , there may well not be M or more individuals who are F times better off than i , benefitting by only a fraction b of individual i 's gain, with no one else affected. Whether all this holds true depends on what the outcomes are and what the size of the population N is. I have therefore framed the Numbers Win property in conditional terms: *If* the only individuals benefiting from y are M or more individuals at level F , etc., then y is ranked better by the SWF despite individual i 's loss.

What happens when \mathbf{U} becomes a non-singleton set? In that case, each $u(\cdot)$ has its own cutoff population M^u . Take some number M which is an upper bound of all these. If there are M or more individuals who are all F times better off than $(x; i)$, benefitting from y by only fraction b of the badly off individual's loss, then the continuous prioritarian SWF would say that y is the better outcome, because it would be better for all $u(\cdot)$ belonging to \mathbf{U} .

Isn't this Numbers Win property very troubling? It brings to mind a famous example provided by Thomas Scanlon, the so-called "transmitter room" example.

Suppose that Jones has suffered an accident in the transmitter room of a television station. Electrical equipment has fallen on his arm, and we cannot rescue him without turning off the transmitter for fifteen minutes. A World Cup match is in progress, watched by many people, and it will not be over for an hour. Jones's injury will not get any worse if we wait, but his hand has been mashed and he is receiving extremely painful electrical shocks. Should we rescue him now or wait until the match is over? Does the right thing to do depend on how many people are watching – whether it is one million or five million or a hundred million? It seems to me that we should not wait, not matter how many viewers there are¹¹¹

In the face of the (seemingly very unattractive) Numbers Win property, how might we defend the continuous prioritarian SWF? As a preliminary matter, it is worth noting that this SWF has a *continuity* property (thus its name). Roughly: if the utility vector associated with x by utility function $u(\cdot)$ is ranked higher than the utility vector associated with y , a sufficiently small perturbation in the first vector will also be better.¹¹²

The only mathematical hitch, here, is that – because \mathbf{U} will have an infinite number of members – the upper bound mentioned in the last paragraph may not exist. However, at least in the focal case where (1) \mathbf{U} is formed as a union of sets measuring each of N spectators' complete extended preferences over life-histories, each consisting of utilities that are unique up to a positive ratio transformation; and (2) the SWF is the Atkinsonian SWF, so the cutoff population is the same for each $u(\cdot)$ and all positive multiples thereof, an upper bound *will* certainly exist. And so, at least in this focal case, the continuous prioritarian SWF does have the unfortunate Numbers Win property.

¹¹¹ Scanlon (1998, 235).

¹¹² Continuity in terms of SWFs is typically framed as follows: a complete ordering of all utility vectors in N dimensional Euclidean space (or of all utility vectors in some orthant of that space, e.g., all utility vectors with nonnegative entries) is continuous iff, for any vector u , the set of vectors that are worse than u , and the set of vectors that are better than u , are each open sets. This, in turn, is true just in case the ordering is representable by a continuous real-valued function.

It is a slight misnomer to say that the SWF now on the table is continuous. The SWF itself is the entire formula: x is at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $\sum_{i=1}^N g(u_i(x)) \geq \sum_{i=1}^N g(u_i(y))$. In describing this

SWF as continuous, what I mean is that, for each assignment of utility vectors to outcomes produced by some $u(\cdot)$, the SWF uses a rule for ordering those vectors which – applied to all utility vectors in N dimensional Euclidean space, or in some orthant -- is continuous. (I add the possibility of the rule ranking only those utility vectors in some orthant because, as we will see, the Atkinsonian SWF malfunctions for negative utilities).

Although continuity in social choice theory is typically discussed with reference to complete orderings of utility vectors, it *is* possible to characterize a quasiordering of utility vectors as continuous, even if the quasiordering is incomplete. See Ok (2007). But, to be clear, I am not suggesting that the quasiordering of *outcomes*, possibly incomplete, produced by what I am calling the "continuous prioritarian SWF" is itself continuous. For starters, it is meaningless to characterize a quasiordering or complete ordering of outcomes as "continuous" or not, absent some metric of distance between the outcomes. Rather, this SWF is "continuous" just in the sense stated above. (Characterizing an ordering of utility vectors as "continuous" *is* possible because a distance metric is available, namely the normal Euclidean distance associated with N -dimensional vectors of real numbers.)

Not only does the SWF using the formula $\sum_{i=1}^N g(u_i(x))$ have this continuity property. It can be shown that *any* prioritarian SWF which has this property yields the same ordering of outcomes as *some* SWF using the $\sum_{i=1}^N g(u_i(x))$ formula.¹¹³

Continuity will appeal to economists as a technically helpful property; and perhaps it has some substantive appeal, as a kind of “stability” property. But, to be honest, it’s hard to see why continuity, alone, makes a moral case for the SWF that incorporates the $\sum_{i=1}^N g(u_i(x))$ formula, as against SWFs which lack the continuity property but also lack the unappealing Numbers Win property.

This section therefore gives closer consideration to the choice between the continuous prioritarian SWF and other prioritarian approaches: in particular, a leximin SWF and a prioritarian SWF with an absolute threshold, both of which lack the Numbers Win property. I argue that, on balance, the continuous prioritarian SWF is more attractive than these alternatives. Although these two particular non-continuous prioritarian SWFs are not the only possible competitors to the continuous prioritarian SWF, these are the only approaches that – as far as I’m aware – have been seriously defended in the literature.

The section closes by discussing the impossibility of constructing an SWF that satisfies all of our intuitive desiderata concerning aggregation and suggesting that Numbers Win, examined more closely, may be less problematic than it seems.

Leximin

The leximin rule for ordering utility vectors has been widely discussed in the SWF literature.¹¹⁴ The idea of a “maximin” approach to choice under uncertainty has long been mooted in decision theory, and was famously relied upon by Rawls in *A Theory of Justice*. A maximin rule for ordering utility vectors would rank them according to the utility of the worst-

¹¹³ Consider any SWF which has the continuity property. In other words, it uses some rule R for ordering utility vectors which – applied to N -dimensional Euclidean space or some orthant -- is continuous; and it says that x is at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $u(x)$ is at least as good as $u(y)$ according to this rule. Because the SWF is prioritarian as well as Paretian, R must be separable, Pigou-Dalton respecting, and Paretian. A basic result in social choice theory is that -- for any continuous separable, Paretian, ordering of utility vectors -- there is some strictly increasing function $h(\cdot)$, such that the additive formula, $\sum_{i=1}^N h(u_i)$, yields the very same ordering. See Blackorby et al. (2005, 116); Bossert & Weymark (2004, 1159). Because we are assuming that the ordering is also Pigou-Dalton respecting, it is not hard to show that $h(\cdot)$ must not only be strictly increasing but also strictly concave. See Adler & Sanchirico (2006, 372).

¹¹⁴ See, e.g., Boadway & Bruce (1984, chapter 5); Bossert & Weymark (2004); d’Aspremont & Gevers (2002); Mongin & d’Aspremont (1998); Blackorby et al (2005, 68-128).

off individual. That rule violates Pareto superiority.¹¹⁵ A leximin rule for ordering utility vectors is a variation on maximin which orders them according to the utility of the worst-off individuals; if these are equal, according to the utility of the second-worst-off individuals; if these are equal, according to the utility of the third-worst-off individuals; and so forth. What I term “the leximin SWF”¹¹⁶ is a generalization of the idea of using a leximin rule to order utility vectors, for the more general context in which we have a set \mathbf{U} of utility functions. The leximin SWF satisfies the basic requirements of Pareto superiority as well as anonymity and Pareto indifference, and is prioritarian.

Leximin lacks the Number Win property, but has another aggregation property which is also unappealing. It prefers to avoid very small losses for a worse off person, even at the expense of foregoing very large benefits for individuals who are only slightly better off. I will call this Absolute Priority for the Worse Off, and will frame it as follows:

Absolute Priority for the Worse Off

In any outcome set, consider any life-history $(x; i)$, however good. Consider any other life-history $(y; i)$, which is worse for the individual i , but perhaps only very slightly so. Choose a positive number M , which can be arbitrarily large; a number F larger than 1, which can be arbitrarily close to 1; and a positive number b , which can be arbitrarily large. Then for an SWF to have the “Absolute Priority for the Worse Off” property means: *Even if* there are M individuals in x , who are better off than i , but the ratio between their well-being and hers is only F ; each of these individuals is better off in y ; their well-being difference is a multiple b of individual i ’s loss; and the only individuals affected by the outcomes are those M individuals and individual i ; *nonetheless* the SWF still ranks x as a better outcome than y .¹¹⁷

¹¹⁵ Imagine two vectors, one of which is Pareto superior, but the worst-off individual has the same utility in both.

¹¹⁶ See above.

¹¹⁷ As with Numbers Win, I have here characterized the well-being level of better off individuals, and their well-being change, as some multiple of the worse-off individuals’, rather than in absolute terms, given limitations to the measurability of well-being.

It is very easy to show that the leximin SWF has the Absolute Priority for the Worse Off property. Consider $(x; i)$ and $(y; i)$. To say that there are M individuals who are only fractionally better off than individual i , by ratio F , is to say: for all $u(\cdot)$ belonging to \mathbf{U} , and for each j in this group, $u(x; j)/u(x; i) = F$. To say that their benefit from y is a b multiple of individual i ’s loss is to say: for all $u(\cdot)$ belonging to \mathbf{U} , and for each j in this group, $(u(y; j) - u(x; j))/(u(x; i) - u(y; i)) = b$.

If, there are M individuals who are F times better than individual i in outcome x , and who benefit from y by a multiple b of his loss (with no one else affected), then – for each $u(\cdot)$ in \mathbf{U} – the leximin rule, applied to $u(x)$ and $u(y)$, says that the first vector is better. Thus the leximin SWF (i.e., x at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $u(x)$ at least as good as $u(y)$ according to the leximin rule) says that outcome x is better. Critically, this is true regardless of the nature of $(x; i)$ and $(y; i)$ – regardless of how well-off individual i would be with $(x; i)$, and how small a difference moving to $(y; i)$ would make to her well-being.

Many authors, observing that the leximin rule gives this sort of absolute priority to the worse off, have rejected it.¹¹⁸

More specifically, I suggest that, as between a continuous prioritarian SWF and leximin, the former is more attractive. First, as between an SWF with the “Numbers Win” property and an SWF with the “Absolute Priority for the Worse Off” property, I suggest that the first type of SWF is, on balance, more plausible. Where the first type of SWF imposes a large loss on a badly off person to confer arbitrarily small benefits on arbitrarily well off individuals, we can always take a stab at justifying this result by pointing to the numbers: it is the fact that a sufficiently large number of people receive a benefit (a small benefit, but a genuine, nonzero benefit nonetheless) that warrants this loss for the badly off person. In the case of the second type of SWF: What justifies foregoing an arbitrarily large benefit for arbitrarily many people? Not necessarily the fact that someone who is badly off stands to lose (because the individual who stands to lose may be quite well off); nor that she is much worse off than those who stand to gain (because the ratio between their well-being and hers may be arbitrarily close to 1); nor the fact that she stands to lose a lot (because the SWF will choose to avoid her loss regardless of what it happens to be).

Second, and perhaps even more importantly, continuous prioritarian SWFs – those that use the $\sum_{i=1}^N g(u_i(x))$ formula – are an entire family of SWFs. The limiting points of this family are, at one end, the utilitarian SWF and, at the other, the leximin SWF. By choosing a sufficiently low degree of concavity for the $g(\cdot)$ function, we can bring these SWFs arbitrarily close to the utilitarian SWF; by choosing a sufficiently *high* degree of concavity for the $g(\cdot)$ function, we can bring these SWFs arbitrarily close to the leximin SWF. In particular, if our SWF is the Atkinsonian SWF, we can bring the SWF indefinitely close to the leximin SWF by increasing the inequality aversion parameter, γ .¹¹⁹

Thus we can always allay our concerns that a particular $g(\cdot)$ function is willing to impose a given loss on a badly off individual so as to benefit a certain number of individuals who are F times better off, whose benefit would only be a fraction b of the loss, by shifting to a $g(\cdot)$ function which is closer to leximin – and would require *more* individuals to benefit by that amount if that loss is to be imposed on the badly off one. Adopting the general functional form, $\sum_{i=1}^N g(u_i(x))$, gives us flexibility to demand that the numbers required to justify a loss for a worse off individual, so as to produce a slight benefit for much better off persons, be arbitrarily large.

¹¹⁸ See, e.g., Arneson (2000, 58); Crisp (2003, 752); McKerlie (1994, 33); Parfit (1991, 121); Temkin (2003b, 78-84); Weirich (1983, 429-30).

¹¹⁹ See Bosmans (2007); Lambert (2001, 97-102)

There is no such flexibility in mitigating the absolutism of the leximin SWF: it is a single SWF, which lies at the limit of the $\sum_{i=1}^N g(u_i(x))$ family.

With these observations in mind, what can be said in defense of the leximin SWF? Interestingly, a number of different axiomatic characterizations of the leximin rule have emerged in the literature that might be used to defend the leximin SWF against continuous prioritarian SWFs and others. One body of literature shows that, if we require an ordering of utility vectors to satisfy certain tradeoff rules in two-person cases – rules that are consistent with, but different from, the Pigou-Dalton axiom -- the leximin rule results. The seminal result, here, is that of Peter Hammond, who shows the following. Imagine that we require an ordering of utility vectors to satisfy the so-called **Hammond equity** condition: Given two outcomes x and y and a utility function $u(\cdot)$: if only two individuals are affected, i and j , with individual i worse off than j in both outcomes; and if individual i is better off in outcome x , and individual j in outcome y ; then x is the better outcome according to $u(\cdot)$, regardless of the difference between individual i 's utility in x and y , and regardless of the difference between individual j 's utility in y and x .¹²⁰

It turns out that leximin is the only rule for ordering utility vectors which is Paretian and anonymous and satisfies the Hammond equity condition.

In important recent work, Bertil Tungodden has shown that an ordering of utility vectors which is Paretian, anonymous and satisfies either of the following conditions also must be the leximin rule¹²¹:

Conditional Contracting Extremes: Given two outcomes x and y and some utility function $u(\cdot)$, if only two individuals are affected, with individual i worse off than j in both outcomes; individual i is better off in x , individual j is better off in y ; individual i is the worst off person in the entire population in both outcomes, and individual j is the best off person in the entire population in both outcomes; *then* outcome x is better than outcome y according to $u(\cdot)$, regardless of the difference between individual i 's utility in x and y , and regardless of the difference between individual j 's utility in y and x . (This implies leximin if we require that the ordering be separable as well as Paretian and anonymous).

Absolute Priority below the Mean: Given two outcomes x and y and some utility function $u(\cdot)$, if only two individuals are affected, with individual i worse off than j in

¹²⁰ See Hammond (1976); Bossert & Weymark (2004, 1151)

Strictly, Hammond equity, as well as the other conditions to be discussed in a moment, are discussed in terms of an ordering of all utility vectors in N -dimensional Euclidean space, or of some orthant, without outcomes being mentioned. So we should add the requirement that the ordering of $u(x)$ and $u(y)$ be generated from a single ordering of N -dimensional Euclidean space (or of some orthant) – which is in fact the approach of all the SWFs discussed in this chapter.

¹²¹ Tungodden (2000). See also Tungodden (2003); Vallentyne (2000); Tungodden & Vallentyne (2005).

both outcomes; individual i is better off in x , individual j is better off in y ; individual i is below the mean utility in y (not necessarily in x); individual j is above the mean utility in both outcomes; *then* outcome x is better than outcome y according to $u(\cdot)$, regardless of the difference between individual i 's utility in x and y , and regardless of the difference between individual j 's utility in y and x .

Each of these equity conditions (Hammond equity, conditional contracting extremes, absolute priority below the mean) can be readily “translated” into the across-outcome conception – namely, into a particular specification of the *strength* of individuals’ claims. For example, the Hammond equity condition in “claim” language says: if individual i has a claim in favor of outcome x , and individual j has a claim in favor of outcome y , and individual i is worse off than j in both outcomes, then i 's claim to x is stronger than j 's claim to y .¹²²

A different axiomatic route from the across-outcome conception of fairness to leximin has also been sketched by Tungodden.¹²³ Imagine that we stick with the basic Pigou-Dalton condition for determining the strength of individuals’ claims. That is, we say: if individual i has a claim in favor of x , individual j in favor of y , then if i is worse off in both outcomes *and* the difference between i 's well-being in the two outcomes is the same as the difference between j 's, i 's claim to x is stronger than j 's claim to y . What Tungodden has shown is that, if we couple this rule for determining the strength of individuals’ claims with a *pairwise aggregation* procedure for adjudicating between conflicting claims where more than two people are affected, plus a requirement that the rankings of outcome sets with different total population sizes be consistent with each other in a certain way, the leximin rule results.

Presumably any plausible specification of the across-outcome conception will say this: Given two outcomes, x and y , if i 's claim to x is stronger than j 's claim to y , and everyone else is unaffected, outcome x is the fairer outcome. (What would it mean for one claim to be “stronger” than another and yet for this condition to fail?) The pairwise aggregation procedure goes much further. It says: If i has a claim in favor of x , and i 's claim in favor of x is stronger than anyone's claim to y , then x is fairer.

Does seeing how the leximin SWF flows from these specifications of across-outcome fairness strengthen the case for leximin? To my mind, it does not. I do not think that Hammond equity, conditional contracting extremes, or absolute priority below the mean states an independently plausible view of the strength of individuals’ claims. Nor do I see a strong reason for thinking that pairwise aggregation is the most attractive rule for adjudicating between

¹²² To translate “conditional contracting extremes” into claims language, simply take the Hammond equity translation and add the condition that individual i be the worst off person in both outcomes, and individual j the best off. Similarly, for “absolute priority below the mean,” add the condition that i be below the mean in outcome y and that j be above the mean in both outcomes.

¹²³ See Tungodden (2003).

individuals' claims where more than two individuals are affected (however we calibrate the strength of claims). To see why pairwise aggregation is problematic, imagine that we (very plausibly) reject the Hammond equity condition for calibrating the strength of claims. If we reject Hammond equity, it is possible for j to have a stronger claim in favor of y than some other individual does in favor of x , even though j is better off than that individual in both outcomes. (Imagine a case in which j stands to benefit a lot and the other individual only a little bit.) Imagine, now, that there are arbitrarily many such individuals, all of whom are worse off than j , and each of whom has a weaker claim to x than j has to y . Regardless of the number of worse-off individuals, pairwise aggregation would force us to count y as fairer, just because a better-off person has a stronger claim than any one of the worse-off individuals. This seems implausible.

Interestingly, Nagel elaborates the idea of pairwise aggregation, but ultimately backs from fully endorsing that approach, recognizing the implausibility of leximin. In his Tanner Lecture he observes:

It seems to me that no plausible theory can avoid the relevance of numbers entirely. If the choice is between preventing severe hardship for some who are very poor and deprived, and preventing less severe but still substantial hardship for those who are better off but still struggling for subsistence, then it is very difficult for me to believe that numbers do not count, and that priority of urgency goes to the worse off however many there are of the better off.¹²⁴

And in his book *Equality and Partiality*: “I am inclined to a somewhat weaker preference for the worse off [than leximin], which can be outweighed by sufficiently large benefits to sufficiently large numbers of those better off.”¹²⁵

In conclusion, I should mention the possibility of a number of additional routes to justifying a leximin SWF that are quite distinct from those we have been analyzing; (1) a *measure-theoretic* argument, having to do with limitations in our ability to measure well-being¹²⁶; (2) a *Rawlsian* route, which sees the moral ranking of outcomes as corresponding to

¹²⁴ Nagel (1977; 125).

¹²⁵ Nagel (1991; 73).

¹²⁶ (1) *Measure Theoretic Arguments*: As mentioned earlier, the theoretical literature on SWFs investigates which rules for ordering utility vectors are invariant to different types of transformation of the utility function. For some such invariance requirements, leximin is the sole prioritarian SWF that satisfies the requirement. In particular, if we require that a complete ordering of utility vectors be Paretian, anonymous and separable, and be invariant to an ordinal transformation (where $u(\cdot)$ is replaced by $\phi(u(\cdot))$, with $\phi(\cdot)$ any strictly increasing function), the only possibilities are leximin and leximax – and leximax is of course not prioritarian (it fails Pigou-Dalton). And, as mentioned earlier, if we require that a complete ordering be Paretian, anonymous, separable, and either utilitarianism or a distribution-sensitive SWF, plus satisfy a requirement of invariance to a positive affine transformation, then (roughly) the only possibilities are leximin or utilitarianism. See Bossert & Weymark (2004; 1154-59).

However, the invariance requirements that would point us to leximin are too strong. Because the methodology for constructing \mathbf{U} makes it meaningful to talk about differences between life-histories, and to compare life-histories to the zero point of nonexistence, we need *not* require invariance to ordinal but non-affine transformations that do not preserve difference comparisons; and we need *not* require invariance to positive affine transformations that change whether life-histories are above or below the zero point. If we require only that the ordering be invariant to a positive ratio transformation, other prioritarian SWFs besides leximin become possible – in particular, the Atkinsonian SWF.

nonprobabilistic self-interested choice behind a veil of ignorance; and (3) the *egalitarian-equivalent* argument, discussed in Chapter 2. For reasons reviewed in the margin, I do not believe that any of these arguments for leximin is successful.¹²⁷

The Prioritarian SWF with an absolute threshold

Campbell Brown has developed the idea of a prioritarian rule for ranking utility vectors that incorporates an absolute threshold – an idea also discussed by Bertil Tungodden.¹²⁸ Brown’s SWF, in our setup, works as follows.

For a given outcome set, identify a life-history, $(x^*; i^*)$ that lies at the “compassion” threshold. For each $u(\cdot)$ belonging to \mathbf{U} , set a utility threshold (for that function) equal to the utility of the threshold life-history. Now compare each pair of outcomes, x and y , using the following two-stage approach. Take some strictly increasing, strictly concave $g(\cdot)$ function. For each utility function $u(\cdot)$ in \mathbf{U} , each outcome corresponds to a below-threshold utility vector,

(2) *The Rawlsian Route.* Imagine that (a) for simplicity, all individuals have identical idealized, self-interested extended preferences over life-histories, lotteries, and comparisons to nonexistence; (b) for an individual to take a moral perspective on outcomes is to determine which one she ideally self-interestedly prefers behind a *nonprobabilistic* veil of ignorance, i.e., seeing each outcome as a package of all its life-histories, without exogenous probabilities; (c) the rational norm for choice without exogenous probabilities is a leximin-type rule (i.e., as between two choices, choose the one whose worst outcome is better; and if the worst outcomes of both choices are equally good, then the second worst; and so forth). Then the upshot would be that the well-being associated with any outcome can be represented by a single utility vector, unique up to a ratio transform, i.e., x is mapped onto $(cu(x; 1), cu(x; 2), \dots, cu(x; N))$; and that each individual’s moral ranking of outcomes would follow a leximin rule.

But both premises (b) and (c) are problematic. Why assume that the moral perspective on outcomes is just a self-interested perspective behind some sort of veil of ignorance, probabilistic or nonprobabilistic? See above. And why think that leximin or maximin or some such nonprobabilistic rule for choice under uncertainty is the appropriate rule, absent exogenous probabilities? Orthodox decision theory says that unique subjective probabilities are always available for the rational chooser, who should maximize expected utility using these. Plausible non-orthodox approaches have been recently developed (for example, the Gilboa/Schmeidler maximin EU approach), which allow the rational chooser to have indeterminate rather than unique and determinate subjective probabilities when exogenous probabilities are not available. But even these approaches say that the rational chooser should be sensitive to her sense of the range of possible probabilities, rather than ignoring probabilities completely – as does a leximin or maximin rule for choice under uncertainty. For more about non-orthodox approaches to choice, see Chapter 7. On maximin and variants (such as leximin) as rules for choice under uncertainty, see [cite to decision theory literature]; Barbera & Jackson (1988).

(3) *Egalitarian Equivalence.* As discussed in Chapter 2, it is possible to develop a welfarist rule for ordering outcomes (in particular, outcomes characterized as bundles of commodities for each individual) by fixing a reference bundle, and giving a higher ranking to x rather than y if the two are Pareto noncomparable and x is “egalitarian equivalent” with respect to the reference bundle, while y is not. For welfarists who care about fairness but are suspicious of interpersonal comparisons, this approach is appealing – because its implementation requires only that we know each individual’s preferences over bundles. And, as discussed, the “egalitarian equivalent” rule is the same as a kind of leximin approach.

However, the premise of this chapter is that utility numbers representing inter- as well as intrapersonal level, difference, and ratio comparisons *are* available to the welfarist. If one accepts this premise, and finds leximin implausible because of the absolute priority it gives to the worse off, why stick with the “egalitarian equivalent” rule for ordering outcomes, as opposed to the continuous prioritarian SWF or some other possibility?

¹²⁷ See also Mariotti & Veneziani (2009), presenting a novel justification for leximin resting on a principle of “noninterference.”

¹²⁸ See Brown (2005, chapter 5; 2005b); Tungodden (2003, 23-29; 2000, 239-42).

namely the vector of individual utilities truncated at the level of the utility threshold. Each outcome also corresponds to an above-threshold utility vector, namely the vector of individual utilities or the threshold utility, whichever is larger. For each utility function, assign outcomes a primary score by summing the elements of the below-threshold vector, transformed by the $g(\cdot)$ function. Then assign outcomes a secondary score by summing the elements of the above-threshold vector, transformed by the $g(\cdot)$ function. If one outcome has a higher primary score it is better according to $u(\cdot)$; if two outcomes have the same primary score and one has a higher secondary score, it is better according to $u(\cdot)$; otherwise the two outcomes are equally good according to $u(\cdot)$. Finally: an outcome is at least as good as outcome y , full stop, iff it is at least as good for all $u(\cdot)$ belonging to \mathbf{U} .¹²⁹

This SWF is identical to the sufficientist SWF, save for the change that the comparison of above-threshold utility vectors is undertaken by summing their elements transformed by the $g(\cdot)$ function, rather than by straight unweighted summation. By virtue of this crucial change, the SWF fully satisfies the Pigou-Dalton principle rather than violating it for transfers among above-threshold individuals. The SWF is therefore prioritarian. Note also that the approach lacks the Numbers Win property which many see as a problematic feature of the continuous prioritarian SWF: in cases where the losing individual is below the threshold, and the individuals who stand to benefit are above, no benefit to them (however large, and however numerous they may be) will counterbalance the loss. Note also that the SWF lacks the Absolute Priority to the Worse Off property which is a problematic feature of leximin: in cases where all the individuals involved are on the same side of the threshold, tradeoffs are allowed.

So the prioritarian SWF with an absolute threshold may seem very attractive. The key difficulty is identifying that threshold. Note that the threshold would not only prevent large losses from being imposed on a below-threshold individual for the sake of very small gains to a sufficiently large number of above-threshold individuals. It would also have the less appealing property of foregoing any gains (even arbitrarily large ones) for individuals above the threshold (however large the group) for the sake of preventing an arbitrarily small loss to a below-

¹²⁹The formalization of this SWF is exactly the same as for the sufficientist SWF, see supra ____, except that the rule for determining whether outcome x is better than y according to $u(\cdot)$ becomes the following: Outcome x is better

than y according to $u(\cdot)$ iff (1) $\sum_{i=1}^N g(u_i^+(x)) > \sum_{i=1}^N g(u_i^+(y))$ or (2) $\sum_{i=1}^N g(u_i^+(x)) = \sum_{i=1}^N g(u_i^+(y))$ and

$\sum_{i=1}^N g(u_i^{++}(x)) > \sum_{i=1}^N g(u_i^{++}(y))$. Outcome x and y are equally good according to $u(\cdot)$ if neither is better than the other.

Similarly, the translation of this SWF into “claim” language is exactly the same as one translation of the sufficientist SWF, which sees individuals’ claims as having a primary and secondary measure (see footnote 102), except for the following change. Each individual i ’s claim in favor of x is measured by a pair of numbers, $(g(u^+(x; i)) - g(u^+(y; i)), g(u^{++}(x; i)) - g(u^{++}(y; i)))$, where $u^+(_ ; i)$ is the i th element of the u^+ vector and $u^{++}(_ ; i)$ is the i th element of the u^{++} vector.

threshold individual.¹³⁰ Is it plausible that morality includes a cut-off like this? Where would it be?¹³¹

The vast literature on poverty metrics, discussed in Chapter 2, is not particularly helpful, here. Poverty metrics, of course, do include a threshold. But these metrics are not, in practice, used to measure the amount of poverty in the distribution of *utility*. Rather, they are used to measure the amount of poverty in the distribution of *income* or (in the case of multidimensional poverty measurement) the amount of poverty as a function of the distribution of various other sources of well-being, such as health, shelter, nutrition, and so forth. The thresholds employed in this literature are therefore thresholds with respect to the sources of well-being (income, health, shelter, etc.), rather than thresholds with respect to well-being itself. One might think of these as discontinuities of some sort *within* the utility function that maps a life-history, described in terms of various sources of well-being, onto a utility number. Moreover, these standard poverty thresholds are set at the level of meeting basic needs – not at some higher point. How shall we move from these source-based and need-focused thresholds, to the moral threshold that the prioritarian SWF now under consideration requires?

It might be thought that the threshold life-history is one in which all the individual's basic needs are met. But then we will be willing to bring an arbitrarily large number of individuals all the way down to the level of basic needs, for the sake of conferring an arbitrarily small benefit on a needy person. As we have seen, Crisp, in specifying his threshold for purposes of the sufficientist SWF, rejects a need-based approach and sets it at the much higher level of a “sufficiently good” life: “eighty years of high quality life on this planet.” This seems totally arbitrarily – and would be equally arbitrary if used as the threshold for purposes of the *prioritarian* SWF now under consideration. Whatever the plausibility of the thought that there is a discontinuity *within* the utility function, at the level of needs, is it plausible that there is a discontinuity at higher levels of well-being sources? If not, what rationale can Crisp give for “eighty years of high quality life”? Note further that a threshold well above the level where all of an individual's basic needs are met will have the unpleasant consequence of foregoing benefits to an arbitrarily large number of non-needy individuals, even for the sake of an arbitrarily small benefit to a non-needy individual below the threshold.¹³²

Numbers Win Redux

¹³⁰Crisp, in his presentation of sufficientism, stipulates that below-threshold individuals take priority over above-threshold individuals with respect to “non-trivial” changes in the well-being of the below-threshold individuals. (2003). It is not clear how to include this caveat in an SWF, and Brown's formalization of the prioritarian SWF with an absolute threshold (and of the sufficientist SWF) does not do so. Even if we *were* to include a “nontriviality” rider in the prioritarian SWF with an absolutist threshold, it would still give absolute priority to any nontrivial change in the well-being of a single below-threshold individual, regardless of how many above the threshold are affected -- and it is not clear how to identify a threshold which would have this sort of force.

¹³¹ See Arneson (2000); Casal (2007).

¹³² Interestingly, Crisp in more recent writing seems to give up on the idea of an absolute threshold. See Crisp (2006; 158).

I have argued that, on balance, continuous prioritarian SWFs are more attractive than the two prioritarian competitors suggested by the existing literature – leximin and a prioritarian SWF with an absolute threshold – despite the fact that continuous prioritarian SWFs possess the “Numbers Win” property. Even if my argument on this score is persuasive, mightn’t some other SWF be better than all of these?

Perhaps so. I’m not sure how I could show that continuous prioritarian SWFs dominate all possible prioritarian competitors, and won’t try to do so here.

However, it can be observed that no prioritarian SWF will possess all the “tradeoff” properties we might find intuitively appealing. On the one hand, it is intuitive that any SWF should be unwilling to impose a substantial loss on a badly off person for the sake of sufficiently trivial gains for individuals who are sufficiently better off, no matter how numerous those individuals. (This is the idea I have tried to capture with the Numbers Win property: intuitively, an SWF should lack that property.) On the other hand, it is intuitive that an SWF should be willing to impose a loss on one person, if sufficiently trivial, where the individuals who benefit are sufficiently numerous. Fleurbaey, Tungodden, and Vallentyne have recently investigated the possibility of crafting moral rules that satisfy both of these desiderata, and have reached negative conclusions.¹³³

In particular, consider what might be termed the “Small Losses Trumped” property:

Small Losses Trumped:

Consider any life-history $(x; i)$ and any other life-history $(y; i)$, in which individual i is worse off. Then there is some positive number l , which is less than 1 and can be arbitrarily close to zero, and which has the following feature. For any pair of outcomes z and z^* that are both better for i than outcome y , if the only individuals affected by this pair of outcomes are (1) individual i , who is worse off in z^* than z , but whose loss is only a fraction l of the well-being difference between $(x; i)$ and $(y; i)$, and (2) other individuals each of whom is better off in z^* than z by the same amount (whatever it may be) and each of whom is better off than i in z , then z^* is a better outcome than z if the number of such individuals is sufficiently large.

The idea here is that, for any well-being loss for a person, there should be some fraction of that loss (perhaps very small) which can be trumped by benefits to a sufficiently large number of people. But it can be shown that any SWF with this desirable property will have the Numbers Win property (at least if we add some technical assumptions.)¹³⁴ This insight, perhaps, will help allay the skepticism of the reader who remains dubious about continuous prioritarian SWFs.

¹³³ See Fleurbaey et al (2008); Fleurbaey & Tungodden (2007).

¹³⁴ Assume that utility is unique up to a positive ratio transformation. (This is the first technical assumption.) Imagine, now, that an SWF has the “Small Losses Trumped” property. Consider life-histories $(x; i)$ and $(y; i)$, such

From Continuous Prioritarian SWFs to the Atkinsonian SWF

Atkinsonian SWFs are a particular subfamily within the broader family of continuous prioritarian SWFs.¹³⁵ As mentioned, they take the form: x is morally at least as good as y iff, for all $u(\cdot)$ belonging to \mathbf{U} , $\frac{1}{1-\gamma} \sum_{i=1}^N u_i(x)^{1-\gamma} \geq \frac{1}{1-\gamma} \sum_{i=1}^N u_i(y)^{1-\gamma}$. SWFs of this sort are sometimes referred to as constant-elasticity-of substitution SWFs, and are widely used in existing scholarly work in the SWF tradition. The inequality aversion parameter, γ , can take any positive value. With $\gamma = 0$ rather than positive, the SWF becomes utilitarian and is no longer prioritarian. As γ increases, the SWF becomes increasingly inequality-averse and, at the limit, approaches leximin. (In the special case where $\gamma = 1$, the Atkinsonian formula is $w(u(x)) = \sum_{i=1}^N \ln(u_i(x))$). How to set γ will be discussed in greater detail, in Chapter 6.

Because they are uniquely defined by a single parameter, γ , Atkinsonian SWFs are mathematically quite convenient. But what really justifies using an Atkinsonian SWF rather than some other type of continuous prioritarian SWF? The answer is measure-theoretic. Atkinsonian SWFs are the only continuous prioritarian SWFs that are invariant to a positive ratio transformation of utility (assuming nonnegative utilities).¹³⁶ Imagine that we impose the following invariance requirement: If the formula $\sum_{i=1}^N g(u_i(x))$ ¹³⁷ yields a particular ordering of outcomes, for a given utility function $u(\cdot)$, then it must yield the same ordering if $u(\cdot)$ is replaced

that the latter is worse for i . Because utility is unique up to a positive ratio transformation, this well-being change can be assigned a utility measure $c\Delta u$, c a positive number. Without loss of generality, assume that $1/l = L$ is an integer.

The Numbers Win property says this: Given life-histories $(x; i)$ and $(y; i)$, if the only individuals affected by outcomes x and y are individual i , and individuals who are F times better off than i in x , and benefit by only a fraction b of individual i 's loss from y , then y is better if those individuals are sufficiently numerous. Imagine that, indeed, there is some number M of individuals in x who are F times better off than i . Imagine, further, that there is a sequence of outcomes $x, z_1, z_2, \dots, z_{L-1}, y$, such that: (1) every one other than the M individuals and i is equally well off in all the outcomes; (2) for each pairing of an outcome in this sequence and the one that succeeds it (i.e., as between x and z_1 , z_1 and z_2 , ..., z_{L-1} and y), individual i is worse off in the second outcome by amount $l(c\Delta u)$, while each of the M individuals is better off by amount $l(c\Delta u)b$. Then Small Losses Trumped entails that, if M is sufficiently large, x is worse than z_1 , z_1 worse than z_2 , ..., z_{L-1} worse than y . Transitivity of course then entails that x is worse than y , which means that the SWF has the Numbers Win property.

The second technical assumption, here, is an outcome set "richness" assumption: namely, that whenever there are some number of individuals at a stipulated multiple F of individual i 's well-being, benefitting by a stipulated amount b from y , then there will also be a sequence of outcome $x, z_1, z_2, \dots, z_{L-1}, y$ as just described. If not, an SWF could fail Numbers Win (preferring x to y regardless of how large M is), without our being able to show that it also fails Small Losses Trumped.

¹³⁵ I refer to them as "Atkinsonian" because they were introduced by Anthony Atkinson in his pioneering work on the use of SWFs to generate an inequality metric.

¹³⁶ See Roberts (1980, 432); Bossert & Weymark (2004, 1161-64); Boadway & Bruce (1984, 159-60).

¹³⁷ As throughout when I use this formula, the $g(\cdot)$ function is strictly increasing and strictly concave.

by $cu(\cdot)$, where c is any positive number. The Atkinsonian formula is the only one that satisfies this requirement.

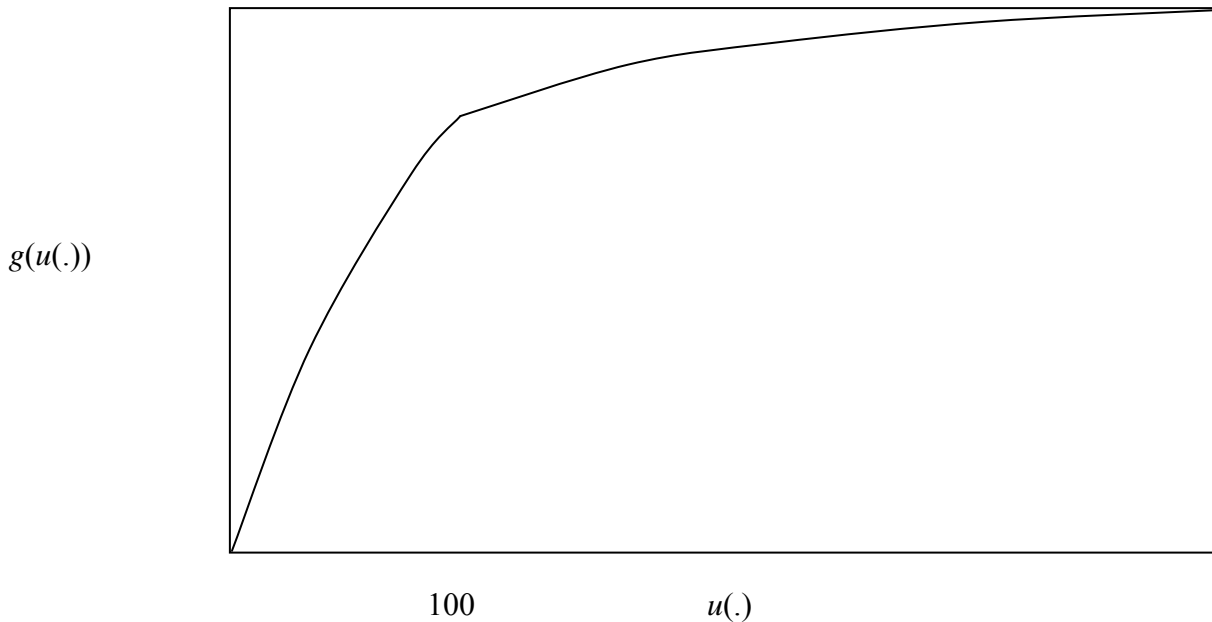
Why require invariance to a positive ratio transformation? Let us back up a bit. In constructing the set \mathbf{U} , I assumed to begin that there may be a legitimate difference between a utility function $u(\cdot)$ and another function $v(\cdot)$ which is an ordinal, but non-affine¹³⁸ transformation of $u(\cdot)$. It may be appropriate to include $u(\cdot)$ in \mathbf{U} but not $v(\cdot)$. Although $u(\cdot)$ and $v(\cdot)$ imply the very same ranking of life-histories, they can imply divergent rankings of differences between life-histories. Well-being differences are a genuine aspect of the concept of well-being, most attractively construed -- and so it may well be appropriate to include $u(\cdot)$ but not $v(\cdot)$ in \mathbf{U} .

I then assumed that well-being ratios are also meaningful – and that nonexistence is the correct zero point for constructing such ratios. Although prioritarrians agree about the Pigou-Dalton principle, they can disagree about “leaky transfers”: about the propriety of a given well-being loss for some individual j , with a smaller gain for an individual i who is worse off than j . The propriety of such a transfer will depend (I assume) on how many times better off j is than i . For example, transferring a small amount of well-being from j , for the sake of individual i benefitting by 1/10th that amount, may be appropriate if j is 10 times better off than j , but not if j is simply twice as well off as i . And – I assume – the correct zero point for ascribing such ratios is nonexistence. Life history $(x; j)$ is twice as good as life-history $(x; i)$, in the sense relevant to determining the propriety of a leaky transfer from j to i , iff the difference between the well-being of the first life-history and nonexistence is the twice the difference between the well-being of the second life-history and nonexistence.

Imagine, now, that $u(\cdot)$ belongs to \mathbf{U} . It yields a particular ranking of life-histories and differences between them, and also assigns zero to nonexistence. Consider, now, $cu(\cdot)$, with c positive. The function $cu(\cdot)$ yields the very same ranking of life-histories and differences between them, and also assigns zero to nonexistence (so that it ascribes the very same ratios between life-histories as $u(\cdot)$). Imagine, now, that an SWF incorporates a rule for ordering utility vectors that yields one ordering of outcomes when $u(\cdot)$ is employed, and a different ordering when $cu(\cdot)$ is employed. Isn't that arbitrary?

For example, imagine a continuous prioritarian SWF with a concave $g(\cdot)$ function that looks like this.

¹³⁸ An affine transformation of $u(\cdot)$ takes the form $au(\cdot) + b$, a and b constants. A positive affine transformation uses a positive a value. An ordinal transformation is any strictly increasing transformation. These include, but are hardly limited to, positive affine transformations.



Such an SWF has a “kink” at 100. Such an SWF might rank the utility vector (90, 40) above (80, 45), but rank the vector (270, 120) below (240, 135) -- even though the second pair of vectors is simply the first pair scaled up by a factor of three – because the SWF treats the fact that an individual’s utility is above or below the level of 100 as meaningful. But what does it mean to say that the utility of a life-history is really 90 rather than 270? What genuine information about well-being is tracked by the numerical fact that the utility attached to a life-history is above or below 100?

Unfortunately, the Atkinson SWF has a serious limitation. It is guaranteed to be invariant to a positive ratio transformation only in the case of utility vectors with nonnegative utilities. It turns out that *no* continuous prioritarian SWF is invariant to a positive ratio transformation when utilities can take on any values, negative or nonnegative.¹³⁹ Further, quite apart from the issue of invariance to a ratio transformation, the Atkinsonian SWF is unattractive if some utilities are

negative. The g -function it uses -- $\frac{u_i(x)^{1-\gamma}}{1-\gamma}$ -- is either undefined or, if defined, not both strictly increasing and strictly concave with negative utilities in the domain of this function.¹⁴⁰

In short, a strong argument can be mounted for using an Atkinsonian SWF to rank an outcome set in which all N individuals have life histories no worse than nonexistence. How to rank outcome sets in which some individuals have life histories worse than nonexistence is a real gap in the prioritarian approach, which this book does not attempt to resolve. The problem, to be

¹³⁹ See Brown (2005; chapter 6); Bossert & Weymark (2004; 1161-64); Blackorby & Donaldson (1982, 258).

¹⁴⁰ In the case of utility vectors that include some zero utilities, the Atkinsonian SWF is undefined for values of γ greater than one, otherwise well-behaved.

clear, is not in assigning a negative utility to a life history, or comparing a life history to nonexistence. The account of well-being developed in Chapter 3 permits such ascriptions without difficulty. The problem arises at the stage of ranking *outcomes*, i.e., combinations of life histories, where some of the life-histories incorporated in outcomes are worse than nonexistence.

Preliminary List of Sources

Adler, Matthew D.; Sanchirico, Chris William. 2006. *Inequality and Uncertainty: Theory and Legal Applications*, 155 **University of Pennsylvania Law Review** 279-377.

Arneson, Richard J. 2000. *Perfectionism and Politics*, 111 **Ethics** 37-63.

Barbera, Salvador; Jackson, Matthew. 1988. *Maximin, Leximin, and the Protective Criterion: Characterizations and Comparisons*, 46 **Journal of Economic Theory** 34-44.

Benbaji, Yitzhak. 2005. *The Doctrine of Sufficiency: A Defence*, 17 **Utilitas** 310-32.

Blackorby, Charles; Bossert, Walter; Donaldson, David. 2005. **Population Issues in Social Choice Theory, Welfare Economics, and Ethics**. (Cambridge, Cambridge University Press).

Blackorby, Charles; Donaldson, David. 1982. *Ratio-Scale and Translation-Scale Full Interpersonal Comparability without Domain Restrictions: Admissible Social-Evaluation Functions*, 23 **International Economic Review** 249-268.

Boadway, Robin; Bruce, Neil. 1984. **Welfare Economics** (Oxford, Basil Blackwell).

Bossert, Walter; Weymark, John A. 2004. *Utility in Social Choice*, in 2 **Handbook of Utility Theory** 1099-1177 (Salvador Barbera et al eds.; Boston, Kluwer)

Bosmans, Kritsof. 2007. *Extreme Inequality Aversion without Separability*, 32 **Economic Theory** 589-594.

Brighouse, Harry; Swift, Adam. 2006. *Equality, Priority, and Positional Goods*, in 116 **Ethics** 471-497.

Brock, Dan W. _____. *Priority to the Worse off in Health-Care Resource Prioritization*.

Broome, John. 1991. **Weighing Goods: Equality, Uncertainty and Time** (Oxford, Basil Blackwell).

Broome, John. _____. *Equality versus Priority: A Useful Distinction* (working paper).

Brown, Campbell. 2003. *Giving up Levelling Down*, 19 **Economics and Philosophy** 111-134.

Brown, Campbell. 2005. *Matters of Priority* (Ph.D. thesis, Australian National University)

- Brown, Campbell. 2005b. *Priority, Sufficiency ... Or Both?* 21 **Economics & Philosophy** 199-220.
- Casal, Paula. 2007. *Why Sufficiency is Not Enough*, 117 **Ethics** 296-326.
- Christiano, Thomas; Braynen, Will. 2008. *Inequality, Injustice and Levelling Down*. 21 **Ratio** (new series) 392-420.
- Crisp, Roger. 2003. *Equality, Priority, and Compassion*, in 113 **Ethics** 745-763.
- Crisp, Roger. 2006. **Reasons and the Good** (Oxford, Oxford University Press).
- D'Aspremont, Claude; Gevers, Louis. 2002. *Social Welfare Functionals and Interpersonl Comparability* in 1 **Handbook of Social Choice and Welfare** 459-541 (Kenneth J. Arrow et al, eds; Amsterdam, Elsevier)
- Doran, Brett. 2001. *Reconsidering the Levelling-down Objection against Egalitarianism*, 13 **Utilitas** 66-85
- Devooght, Kurt. 2003. *Measuring Inequality by Counting "Complaints": Theory and Empirics*, 19 **Economics and Philosophy** 241-263.
- Fleurbaey, Marc. _____. *Equality versus Priority: How Relevant is the Distinction?* (working paper)
- Fleurbaey, Marc; Tungodden, Bertil. 2007. *The Tyranny of Non-Aggregation versus the Tyranny of Aggregation in Social Choices: A Real Dilemma* (working paper).
- Fleurbaey, Marc; Tungodden, Bertil; Vallentyne, Peter. 2008. *On the Possibility of Nonaggregative Priority for the Worst Off*, _ **Social Philosophy & Policy** 258-285.
- Frankfurt, Harry. 1987. *Equality as a Moral Ideal*, 98 **Ethics** 21 (1987).
- Hammond, Peter J. 1976. *Equity, Arrow's Condition, and Rawls' Difference Principle*, 44 **Econometrica** 793-804.
- Hausman, Daniel. _____. *Equality versus Priority: A Badly Misleading Distinction*.
- Holtug, Nils. 2007. *Prioritarianism*, in **Egalitarianism: New Essays on the Nature and Value of Equality** 125-56(Nils Holtug & Kasper Lippert-Rasmussen, eds.; Oxford, Oxford University Press).
- Holtug, Nils. 2007b. *A Note on Conditional Egalitarianism*, 23 **Economics and Philosophy** 45-63.

Jensen, Karsten Klint. 2003. *What is the Difference between (Moderate) Egalitarianism and Prioritarianism?*, 19 **Economics and Philosophy** 89-109.

Jensen, Karsten Klint. __. *Measuring the Size of a Benefit and its Moral Weight: On the Significance of John Broome's "Interpersonal Addition Theorem."*

Lambert, Peter. 2001. **The Distribution and Redistribution of Income** (Manchester, Manchester University Press, 3d ed.)

Lumer, Christoph. 2005. *Prioritarian Welfare Functions: An Elaboration and Justification*, in **Democracy and Welfare** (Daniel Schoch ed., Paderborn, Mentis) [**has this been published?**]

Mariotti, Marco; Veneziani, Roberto. 2009. "Non-interference" Implies Equality, 32 **Social Choice & Welfare** 123-128.

Mason, Andrew. 2006. **Levelling the Playing Field: The Idea of Equal Opportunity and its Place in Egalitarian Thought** (Oxford, Oxford University Press).

Mason, Andrew. 2001. *Egalitarianism and the Levelling Down Objection*, 61 **Analysis** 246-254.

McKerlie, Dennis. 1994. *Equality and Priority*, 6 **Utilitas** 25-42.

Mongin, Philippe; D'Aspremont, Claude. 1998. *Utility Theory and Ethics*, in 1 **Handbook of Utility Theory** 371-481 (Salvador Barbera et al eds.; Dordrecht, Kluwer)

Moreno-Ternero, Juan D.; Romer, John E. 2008. *The Veil of Ignorance Violates Priority*, 24 **Economics & Philosophy** 233-57.

Nagel, Thomas. 1977. *Equality*, in **Mortal Questions** 106-127 (Cambridge, Cambridge University Press, 1979)

Nagel, Thomas, 1991. **Equality and Partiality** (Oxford, Oxford University Press).

O'Neill, Martin. 2008. *What Should Egalitarians Believe?* 36 **Philosophy & Public Affairs** 119-156.

Ok, Efe. 2007. **Real Analysis with Economic Applications** (Princeton: Princeton University Press).

Parfit, Derek. 1991. *Equality or Priority?*, in **The Ideal of Equality** 81-125 (Matthew Clayton & Andrew Williams eds; Houndmills, Macmillan, 2000).

Persson, Ingmar. 2001. *Equality, Priority and Person-Affecting Value*, 4 **Ethical Theory and Moral Practice** 23-39.

- Petererson, Martin; Hansson, Sven Ove. 2005. *Equality and Priority*, 17 **Utilitas** 298-309.
- Rabinowicz, Wlodek. 2001. *Prioritarianism and Uncertainty: On the Interpersonal Addition Theorem and the Priority View*, in **Exploring Practical Philosophy: From Action to Values** 139-____ (____Egonsson et al eds; _____)
- Ramsay, Marc. 2005. *Teleological Egalitarianism versus the Slogan*, 17 **Utilitas** 93-116 .
- Roberts, Kevin W.S. 1980. *Interpersonal Comparability and Social Choice Theory*, 47 **Review of Economic Studies** 421-439.
- Scanlon, T.M. 1998. **What We Owe to Each Other**. (Cambridge, Harvard University Press).
- Temkin, Larry S. 1993. **Inequality** (Oxford, Oxford University Press).
- Temkin, Larry S. 2000. *Equality, Priority, and the Levelling Down Objection*, in **The Ideal of Equality** 126-161 (Matthew Clayton & Andrew Williams eds; Houndmills, Macmillan, 2000).
- Temkin, Larry S. 2003a. *Egalitarianism Defended*, 113 **Ethics** 764-782.
- Temkin, Larry S. 2003b . *Equality, Priority or What?*, 19 **Economics and Philosophy** 61-87.
- Temkin, Larry S. 2003c. *Personal versus Impersonal Principles: Reconsidering the Slogan*, ____ **Theoria** 21
- Temkin, Larry S. 2003d. *Determining the Scope of Egalitarian Concern: A Partial Defence of Complete Lives Egalitarianism*, ____ **Theoria** 46.
- Temkin, Larry S. 2003e. *Measuring Inequality's Badness: Does Size Matter? If So, How, If Not, What Does?*, ____ **Theoria** 85
- Tungodden, Bertil. 2003. *The Value of Equality*, 19 **Economics and Philosophy** 1.
- Tungodden, Bertil; Vallentyne, Peter. *On the Possibility of Paretian Egalitarianism*, 102 **J. Philosophy** 126-154.
- Vallentyne, Peter. 2000. *Equality, Efficiency, and the Priority of the Worse Off*, 16 **Economics and Philosophy** 1-19.
- Weirich, Paul. 1983. *Utility Tempered with Equality*, 17 **Nous** 423-439.
- Weymark, John A. 1991. *A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism*, in **Interpersonal Comparisons of Well-Being** 255-320 (Jon Elster & John E. Roemer eds; Cambridge, Cambridge Univeristy Press)