Matthew D. Adler,1 “Theory of Prioritarianism”

This paper is a draft of Chapter 2 of Matthew D. Adler and Ole Norheim, eds., Prioritarianism in Practice (Cambridge University Press, forthcoming).

December 2020

Abstract

This chapter provides theoretical foundations for the Prioritarianism in Practice volume. It does so by analyzing the features of prioritarian social welfare functions (SWFs). A prioritarian SWF sums up individuals’ well-being numbers plugged into a strictly increasing and strictly concave transformation function. Prioritarian SWFs, like the utilitarian SWF, fall within the “generalized utilitarian” class of SWFs. Generalized-utilitarian SWFs are additive and, hence, especially tractable for purposes of policy analysis. The chapter reviews the axiomatic properties of generalized utilitarian SWFs and, specifically, of prioritarian SWFs. Prioritarianism satisfies the Pigou-Dalton axiom (a pure, gap-diminishing transfer of well-being from a better-off to a worse-off person is an ethical improvement), while utilitarianism does not. Pigou-Dalton is the axiomatic expression of the fact that a prioritarian SWF gives extra weight (priority) to well-being changes affecting worse-off individuals.

The chapter also discusses the informational requirements of prioritarian SWFs (as regards interpersonal well-being comparisons). It reviews the various methodologies for applying a prioritarian SWF under uncertainty. And it describes the two main subfamilies of prioritarian SWFs, namely Atkinson and Kolm-Pollak SWFs.

Keywords: Prioritarianism, utilitarianism, social welfare function (SWF), Pigou-Dalton, Separability, interpersonal comparisons of well-being, generalized-utilitarian SWFs, Atkinson SWFs, Kolm-Pollak SWFs

2.1 Introduction

This chapter is the second chapter in Prioritarianism in Practice. See https://web.law.duke.edu/laweconomicsandpublicpolicy/pip/book/. See Adler and Norheim, chapter 1, for an overview of the book. In a nutshell, Prioritarianism in Practice aspires to advance our understanding of prioritarianism as a policy assessment tool. The book does so by considering the application of prioritarianism to a variety of policy domains, including taxation (Tuomala and Weinzierl, chapter 4); health care (Cookson, Norheim, and Skarda, chapter 6);

1 Richard A. Horvitz Professor of Law and Professor of Economics, Philosophy and Public Policy, Duke University. adler@law.duke.edu. Thanks to Francisco Ferreira, Erwin Ooghe, Ole Norheim, Paolo Piacquadio, and to participants in the Prioritarianism in Practice book conferences at Duke University and University of Bergen, for their comments.
fatality risk regulation (Hammitt and Treich, chapter 7); climate change (Ferranna and Fleurbaey, chapter 8); education (Ooghe, chapter 9); and the Covid-19 pandemic (Bloom, Ferranna, and Sevilla, chapter 12). Prioritarianism can also be used as a methodology for assessing societal condition, analogous to GDP (Decancq and Schokkaert, chapter 5). For all these applications, it is relevant to know whether individuals’ ethical preferences are in fact prioritarian and—if so—what degree of priority is given to the worse off (Schokkaert and Tarroux, chapter 10).

The intention of this chapter is to provide theoretical foundations for the Prioritarianism in Practice enterprise, by reviewing the theory of prioritarianism in some detail. The chapter interlocks with chapter 3 (Adler and Decancq), which covers one major theoretical problem not addressed here—namely, how to construct an interpersonally comparable well-being measure. Chapter 11 (Brunori, Ferreira and Peragine) is also an important theory chapter; it considers how to modify prioritarianism so as to be sensitive not only to the pattern of individual well-being, but also to considerations of individual responsibility.

Utilitarianism makes overall well-being the master criterion for ethics. One outcome is ethically better than a second if and only if (iff) the sum total of individual well-being is larger. Utilitarianism is, of course, very well entrenched both in philosophy and in economics. It dates back hundreds of years to the writings of Jeremy Bentham; has inspired generations of philosophical scholarship; and also lies behind much of the literature in normative (welfare) economics. (Blackorby, Bossert and Donaldson, 2002; Brandt, 1979; Broome, 1991; Eggleston and Miller, 2014; Hare, 1981; Harsanyi, 1977; Mongin and d’Aspremont, 1998; Sen and Williams, 1982; Singer, 2011; Smart and Williams, 1973).

Prioritarianism is a newer entry to the history of ethical thought. It ranks outcomes according to the pattern of well-being, but so as to give priority (hence the term “prioritarianism”) to individuals at lower well-being levels. (As will be discussed below, this is accomplished by summing transformed well-being numbers—more precisely, well-being numbers plugged into a strictly increasing and strictly concave transformation function.) Prioritarianism was introduced to philosophy at the end of the twentieth century, and to economics a few decades earlier. In philosophy, prioritarianism is most closely associated with Derek Parfit—who described the view in lectures given in 1991, calling it the “priority view,” which in turn morphed into the term “prioritarianism,” now canonical among philosophers. (Parfit’s 1991 lecture was published in Parfit [2000].)³ For overviews of the philosophical literature on prioritarianism, see Adler (2012, ch. 5); Adler and Holtug (2019); Holtug (2010, 2017); Otsuka and Voorhoeve (2018). On the history of prioritarianism (not using that term) in economics, see Adler and Norheim, chapter 1, this volume.

---

² This chapter does address a range of issues regarding well-being, as the reader will see; but it “hands off” the problem of constructing an interpersonally comparable well-being measure to Chapter 3.

³ Philosophers frequently credit Parfit with the invention of prioritarianism, although in fact there is work by a few other philosophers on the concept of prioritarianism (not using that term) that slightly pre dates Parfit’s. (McKerlie 1984; Temkin 1983; Weirich 1983)
Utilitarianism and prioritarianism are both species of welfare consequentialism. (Adler 2012, ch. 1.) According to consequentialism, there is an ethical ranking of outcomes; and the ethical ranking of choices in any given choice set (the set of choices available to some decisionmaker, at a given point in time) is determined by the ranking of outcomes. Welfare consequentialism constrains the outcome ranking so as to depend upon individual well-being. Formally, this is accomplished via the Pareto Indifference axiom: If two outcomes, \( x \) and \( y \), are such that each person is equally well off in \( x \) as she is in \( y \), then \( x \) and \( y \) are equally ethically good. Both utilitarianism and prioritarianism satisfy this axiom.

The social welfare function (SWF) framework is a systematic decision procedure for implementing welfare consequentialism. Governmental decisionmakers who wish to operationalize prioritarianism can most rigorously do so by means of the SWF framework: by employing a prioritarian SWF. This book, *Prioritarianism in Practice*, therefore operates within the SWF framework. It examines the use of prioritarian SWFs as a policy-assessment methodology.

In turn, therefore, the present chapter will set forth the theory of prioritarianism via the SWF framework. The chapter's aims are both explicative and argumentative: to explain how prioritarianism is formalized as an SWF; and to present arguments in defense of a prioritarian SWF rather than alternative functional forms.

Although the topic of the chapter is “Theory of Prioritarianism,” utilitarianism is also discussed in considerable depth. Prioritarianism has many similarities to utilitarianism; yet, in certain critical respects, the two are different. One gains a clearer understanding of prioritarianism by seeing it how it both relates to, and innovates upon, the dominant approach within ethics and welfare economics, namely utilitarianism.

More precisely, both the utilitarian SWF and prioritarian SWFs fall within a broader class of SWFs: generalized-utilitarian SWFs. The argument offered by the chapter in defense of prioritarianism will have a two-part structure: first, a defense of generalized-utilitarian SWFs over competitors (Section 2.4); and second, within the generalized-utilitarian class of SWFs, a defense of prioritarianism over utilitarianism (Section 2.8).

---

4 “Outcomes,” here, might be understood as possible worlds, or alternatively as cognitively tractable models of possible worlds. This chapter takes the latter position. See Section 2.2.1.

5 This chapter focuses on the SWF framework as a decision procedure. However, as already suggested, we might employ prioritarianism not merely as a policy-assessment methodology (that is, as guidance to governmental officials in making policy choices), but also as a tool for evaluating societal condition. This application of the SWF framework is covered in Decancq and Schokkaert, chapter 5, this volume.

6 The chapter presents straight normative (ethical) arguments in favor of prioritarianism. These arguments are certainly contestable, and in particular reach conclusions that may not be well reflected in citizens’ ethical preferences. See Schokkaert and Tarroux, chapter 10, this volume. How the policymaker should take account of these dual inputs into decision—the ethical arguments presented to her by academics or others who engage her in ethical deliberation, and information about citizens’ ethical preferences—is a complicated topic not addressed here. (Adler 2019, ch. 6)
Section 2.2 provides an overview of the SWF framework. This framework revolves around three critical formal tools: a well-being measure \( w(\cdot) \), for converting outcomes into lists ("vectors") of well-being numbers; an SWF, which is a rule for ranking these well-being vectors, abbreviated as \( \succeq \); and a procedure for applying the SWF under uncertainty. It presents formulas for the most widely used SWFs, including the utilitarian SWF, prioritarian SWFs, and others (the leximin SWF, rank-weighted SWFs, and sufficientist SWFs). Utilitarianism is a specific SWF (the simple sum of well-being numbers), while prioritarianism is a family of SWFs (defined by all possible strictly increasing and strictly concave transformation functions).

Section 2.3 discusses the axiomatic characterization of SWFs. The major axioms useful in demarcating the space of SWFs are: Pareto (meaning the combination of Pareto Indifference and Strong Pareto), Anonymity, Continuity, Separability, and Pigou-Dalton. The first two are generally accepted by welfare-consequentialists and, indeed, are satisfied by all major SWFs; this is not true of the remaining three axioms. Generalized-utilitarian SWFs satisfy Pareto, Anonymity, Continuity and Separability. Pigou-Dalton, an equity axiom, is not satisfied by the utilitarian SWF, but is satisfied by prioritarian SWFs.

Section 2.4 offers a defense of the generalized-utilitarian class of SWFs.

Section 2.5 addresses well-being. The SWF framework requires interpersonally comparable well-being numbers. This part of the chapter explains why.

Section 2.6 takes on a question sometimes raised about prioritarianism, namely: is prioritarianism truly distinct from utilitarianism? The answer is “yes.” In a nutshell: the outcome ranking that results from a prioritarian SWF in combination with a given well-being measure \( w(\cdot) \) can be mimicked by conjoining the utilitarian SWF with a different well-being measure \( w^*(\cdot) \). But utilitarianism and prioritarianism are distinct relative to any particular well-being measure \( w(\cdot) \). If \( w(\cdot) \) is the correct well-being measure—correct in the sense of capturing our judgments about intra- and interpersonal comparisons of well-being levels and differences—then \( w^*(\cdot) \) is not.

Section 2.7 discusses the application of prioritarianism and utilitarianism under uncertainty. This topic has great practical relevance (human policymakers operate under uncertainty), but also theoretical import. Utilitarianism under uncertainty satisfies the ex ante Pareto axiom, while prioritarianism under uncertainty runs afoul of that axiom. (To be precise: prioritarianism under uncertainty is unable to satisfy both the ex ante Pareto axiom and a Dominance axiom that captures the essence of consequentialism under uncertainty.)

Section 2.8 engages the normative debate between utilitarianism and prioritarianism.

Section 2.9 discusses the two major types of prioritarian SWFs: Atkinson SWFs and Kolm-Pollak SWFs.
Section 2.10 explains the relation between prioritarian SWFs and inequality metrics.

Section 2.11 discusses variable population. The chapter generally simplifies the exposition of the SWF framework by assuming a fixed population of ethical concern: the very same individuals exist in each outcome. Section 2.11 relaxes this assumption.

The chapter includes an Appendix, which provides a more formal treatment of certain points in the main text.

Because the SWF framework ranks outcomes in light of their corresponding well-being vectors, it is insensitive to facts about outcomes except insofar as these “show up” in individuals’ well-being. Assume that some individuals are better off in outcome \( y \) than outcome \( x \), while others are better off in \( x \) than \( y \). Now assume that the pattern of well-being in \( x^* \) is the same as in \( x \), and the pattern of well-being in \( y^* \) is the same as in \( y \). Then the prioritarian SWF, the utilitarian SWF, and indeed any (anonymous) SWF\(^7\), will rank the first pair of outcomes the same way as the second. But there might be good non-welfare reasons to rank \( x \) better than \( y \) but \( y^* \) better than \( x^* \). In particular, non-welfare facts regarding individual responsibility might argue for a differentiated ranking. (For example, some of the individuals who are at low well-being levels in \( x \) are responsible for their plight; they have knowingly made imprudent choices. In \( x^* \), the counterpart individuals at these low well-being levels lack responsibility; they are the victims of “brute luck.”)\(^8\)

As already mentioned, it is the burden of Chapter 11 to discuss how to generalize prioritarianism to take account of responsibility. That topic is not addressed here.

2.2 The SWF Framework

The SWF framework is a welfare-consequentialist decision-procedure. It is a rigorously structured methodology for implementing welfare-consequentialist constructs such as utilitarianism or prioritarianism.\(^9\)

The framework’s components are as follows: A set of outcomes \( \mathbf{O} = \{x, y, \ldots \} \); a population of individuals of ethical concern, understood for now as a fixed and finite set of individuals \( \mathbf{I} = \{1, \ldots, N\} \),\(^10\) each of whom exists in all of the outcomes; a set of policies (possible governmental actions) \( \mathbf{P} = \{P, P^*, P^{**}, \ldots \} \); a well-being measure \( w(\cdot) \); an SWF,

\(^7\) An “anonymous” SWF is one that satisfies the Anonymity axiom. All the major SWFs (utilitarian, prioritarian, leximin, rank-weighted and sufficientist) do so. If two outcomes have the same well-being pattern (in the sense that the allocation of well-being to the \( N \) individuals in the population in the second outcome is just a permutation of the allocation in the first), then Anonymity requires the two outcomes to be ranked equally good. See Section 2.3.

\(^8\) The importance of individual responsibility as a component of ethical assessment is the core insight in the philosophical literature on “luck egalitarianism,” (Knight, 2009; Lippert-Rasmussen, 2016), which in turn is the inspiration for the economic literature on equality of opportunity (Ferreira and Peragine, 2016).

\(^9\) The SWF framework, as summarized in this section, is set forth in detail in Adler (2019).

\(^10\) \( N \geq 3 \). \( N = 1 \) does not pose a problem of social choice; and with \( N = 2 \) the axiom of Separability is not meaningfully distinct from the Pareto axiom but is rather implied by it (see section 2.3 for statements of axioms).
abbreviated as \( \succcurlyeq \), which is a rule for ranking well-being vectors; and an uncertainty module for the SWF. I’ll discuss each of these in turn.

### 2.2.1 Outcomes, Individuals, Policies

What is an “outcome”? As a first cut, we might say that each outcome is a “possible world.” The consequentialist perspective on ethics says: the ethical ranking of the various choices available to a decisionmaker depends on what might happen as a result of each choice, and on how good those consequences are. A possible world is a complete description of a possible history of the universe. It tells us “what might happen” in full detail. The description is complete in the sense that every possible fact is either included within the description or precluded by it. (So, for example, a possible world will specify either that Matt Adler sneezes at 11 a.m. on December 17, 2019, or that he does not sneeze then.) Thinking about which possible worlds might result from each choice available to us brings into play literally everything we might care about, ethically.

However, human decisionmakers are cognitively bounded. A human being doesn’t have the mental capacity to think about a possible world, described in full detail, let alone a set of possible worlds. In order for “outcomes” to play a role within a decision procedure that a human decisionmaker can actually use, outcomes can’t be possible worlds. Rather, outcomes are cognitively tractable models of possible worlds. Each outcome in the set \( \mathbf{O} \) is a description of a possible combination of certain facts, in particular certain facts of ethical relevance.

For welfare-consequentialists, the facts of ethical relevance are facts about individual well-being. An individual’s well-being depends upon her characteristics or “attributes.” Economists standardly model an individual’s well-being as depending on some or all of the following attributes: income or consumption, leisure, health, longevity, and her level of public goods (e.g., environmental amenities). To be sure, many other types of attributes may be well-being-relevant; which ones are depends upon the theory of well-being (an issue I’ll turn to momentarily). Further, for reasons of cognitive tractability, an outcome can’t describe all of an individual’s welfare-relevant attributes, but only a subset thereof.

Pulling this together: a given outcome \( x \) is a possible series of attribute bundles, one for each member of the population of concern. Let \( a \) denote an attribute bundle, and \( a_i \) the attribute bundle of individual \( i \). Then \( x = (a_1, a_2, \ldots, a_N) \), namely the combination of a possible attribute bundle for individual 1, a possible attribute bundle for individual 2, \ldots, a possible bundle for individual \( N \). A given bundle specifies an individual’s holdings of some of the attributes of well-being relevance: for example, her income and leisure; or her income, leisure, and health; or her income and longevity. Which attributes to include in the bundle description is a modeling

---

11 Because we are interested in lifetime well-being, each attribute bundle is a lifetime sequence of attributes. See Section 2.2.2.
question. To repeat: cognitive tractability means that a bundle can’t include all of the individual’s welfare-relevant characteristics, but only some of them.\footnote{It should be stated that the notion of an attribute bundle is meant to allow for relational attributes (expressing how an individual is situated relative to others, e.g., whether the individual has a below-average income) as well as the non-relational attributes more typically used by economic modelers (e.g., the individual’s income). If a well-being theory counts certain relational attributes as a determinant of an individual’s welfare, those can be included in the bundles. To be sure, if bundles include relational attributes, then the combination of bundles in a given outcome will need to be internally consistent. (For example, it’s not possible for everyone to have below-average income.)}

Actually, this characterization of outcomes isn’t quite adequate for a \textit{preference-based} view of well-being. Those holding such a view give a special role to an individual’s preferences in determining her well-being. They’ll want to differentiate between two types of individual characteristics that are well-being relevant: non-preference characteristics, and preferences themselves. So as to accommodate this view, I’ll use the term “attribute” to indicate some type of non-preference characteristic. An attribute bundle will be abbreviated as $a$, and a preference as $R$. Let a “history” denote a combination of an attribute bundle and a preference. History $h = (a, R)$. On a preference view, a given outcome is a possible series of histories (each history a combination of an attribute bundle and a preference), one history for each member of the population of concern.\footnote{That is: A given outcome $x = ((a_1, R_1), \ldots, (a_n, R_n))$, with $(a_i, R_i)$ the history of individual $i$. Adler and Decancq, chapter 3, this volume, explore how well-being is assigned to histories on the assumption that each $R$ is a single preference; but the “history” concept can be generalized to allow for preference change by letting $R$ be an intertemporal sequence of preferences.}

What is the population of ethical concern? The ethical perspective is \textit{impartial}. Welfare consequentialists, to be fully impartial, should accord equal weight to the well-being of all human beings, at least. Actually, fully impartial welfare consequentialists should give equal weight to the well-being of all humans and also of all non-human animals that (like humans) are “welfare subjects.” However, so as to avoid a variety of difficult normative and implementation questions that arise in extending the SWF framework to animals, this chapter (and indeed the entire book) focuses on the simpler case of a population of ethical concern that includes only humans. “Person” and “individual” will be used as a synonym for “human being.”\footnote{It might be argued that conceptualizing each outcome as \textit{merely} a series of attribute bundles or a series of histories is too restrictive. These welfare-relevant facts about outcomes are sufficient to translate each outcome into a well-being vector, but not for purposes of the SWF’s uncertainty module, whereby a given policy maps states-of-nature onto outcomes. See Section 2.2.4. In order to ascertain that mapping, we will need to know the causally-relevant features of outcomes that determine which outcome would result from given policy $P$ in a given state of nature $s$. This objection is easily handled. We can simply say that each outcome \textit{includes} a list of attribute bundles/histories, one for each individual, and also may include further facts—as needed for the uncertainty module.}

To be precise, I assume that each individual in the population is a human being and a person—that is, a human being with an array of psychological attributes sufficient to make him/her a person. These are attributes such as sentience, having preferences and beliefs, self-consciousness, the capacity for autonomous choice, etc. What these psychological attributes are is contested, in the philosophical literature on personhood; for purposes of this chapter, it’s sufficient to say that the individuals in the population are persons, whatever exactly that means.
Policymakers are rarely fully impartial, or even aspire to be. A governmental decisionmaker might well take as her aim the well-being of citizens; non-citizen interests might be ignored or downweighted. The SWF framework is applicable to any population, whether that population is all humans or a subset—as long as everyone within the population is accorded equal weight.\footnote{Actually, the SWF framework itself doesn’t require equal weighting. What does so is the Anonymity axiom, to be discussed below, Section 2.3, which we’ll take as a foundational axiom.} Although powerful ethical arguments support the position that the population of ethical concern is all humans, not a subset, the SWF methodology can be used in a more restricted way. Further, all of the other issues regarding the specification of the methodology—the well-being measure, functional forms for the SWF, and application under uncertainty—are orthogonal to how the population is defined. Thus, for the remainder of the chapter, “population of ethical concern” or, for short, “population” will be either all humans or a subset (e.g., the citizenry of some country).

The most straightforward application of the SWF framework involves a \textit{fixed} and \textit{finite} population. “Fixed” means that the very same individuals exist in every outcome. \textbf{I}, the population of ethical concern, is such that each member of \textbf{I} exists in every outcome in \textbf{O}.

Various types of beings are \textit{welfare-subjects}—they have a well-being—but are not human persons. See generally Singer (2011). These include: (a) non-human animals that are sentient but are not persons; (b) non-human animals (if there are any) that are persons or close to persons in their psychological attributes; (c) human beings with psychological impairments so serious that they are not persons, but are still sentient and possess a well-being; and (d) non-living beings, such as highly intelligent computers, that are nonetheless persons or at least possess a well-being. Welfare consequentialists ascribe ethical status to all such non-human-person welfare subjects (“NHP” welfare subjects). But extending the SWF framework and, specifically, prioritarianism to include them raises many complications. (1) Can we make well-being comparisons between human persons and NHP welfare subjects? (2) There is a strong case that the inputs to the SWF framework should be lifetime well-being numbers, in virtue of the psychological continuity of human persons over time. See Section 2.2.2. Some NHP welfare subjects lack such continuity. How should the SWF framework be specified to take account of the lifetime well-being of human persons as compared to the sublifetime (momentary?) well-being of such NHP welfare subjects? (3) Assuming we can make well-being comparisons between human persons and NHP welfare subjects, it seems intuitively plausible that many are worse off than human persons. For example, most species of sentient non-human animals would seem to be, on average, worse off than human persons. Applying a prioritarian SWF to all these beings would mean giving priority to the interests of the former over the latter—which seems quite counterintuitive (although this intuition may just reflect a pro-human bias on the part of the author!).

It is beyond the scope of this chapter to address these difficult questions. I simply assume that each individual in the population is a human person. Thus, throughout the text of the chapter, “human,” “individual,” and “person” are used interchangeably as shorthand for “human person.” Further, although in the future the prioritarianism-in-practice project should certainly broaden its agenda to consider NHP welfare subjects, none of the chapters in the current volume do so. For an initial attempt to include non-human animals within the SWF framework, see Johansson-Stenman (2018).

A final note. No adult human being had the characteristics of a person at birth. Rather, any such adult acquired those characteristics at some point during her infancy or childhood. (What time that was depends on the precise list of characteristics sufficient for personhood.) Imagine that Lana, a human being, dies in outcome $x$ at the age of 3 months, before becoming a person. How do we compare Lana’s (lifetime?) well-being in $x$ to that of human persons? Assuming we can make such comparison, it would seem that Lana-in-$x$ is worse off than ordinary human persons, because her life is shorter. Should she, therefore, take priority over them? How the SWF framework should take account of human beings who die before becoming persons is, again, a quite important question but beyond what I can address in this chapter. I’ll circumvent it by assuming that, in each outcome, each human being is old enough to be a full person. (Gamlund and Solberg [2019] grapples with the problem of determining the ethical value of saving newborns as compared with children and adults.)
“Finite” means that the number of individuals in I is some finite number N. As shorthand, I’ll use a distinct number (rather than a distinct proper name) to refer to each member of I. So I = \{1, 2, …, N\}, with individual 1 a particular person, individual 2 another particular person, and so forth. The assumption of a fixed and finite population is adopted throughout the chapter until section 2.11. See note __ for an explanation of how this assumption is consistent with the analysis of policies that affect longevity, as in Cookson, Norheim and Skarda (chapter 6, this volume) and Hammitt and Treich (chapter 7).

The set \( P = \{P, P^*, \ldots\} \) is a formal representation of some decisionmaker’s choice situation. At a given point in time, the decisionmaker is selecting among possible choices—possible actions she might perform. P, the choice set, includes all the actions/choices that the decisionmaker is considering. The SWF framework will help her evaluate these choices. It will do so by ranking them, from best to worst.

I assume, specifically, that the decisionmaker is a governmental official of some type (or a governmental institution), and thus each choice in P is a policy: some course of action by government, such as enacting a particular statute, promulgating a particular regulation, building some particular infrastructure, distributing funds in some particular way, and so forth. (Thus the use of “\( P \)” to denote the set and “\( P \)” a member, as mnemonics for “policy.”) Actually, the SWF framework is equally usable by private individuals and organizations. Consequentialism famously stipulates (notoriously so, for those ethicists who oppose consequentialism) that everyone is under an ethical injunction to promote ethically good outcomes. This is just as true for an ordinary person as for a governmental actor. Consequentialists draw no demarcation between the ethics of private and governmental choice. However, since in practice the SWF methodology has been, and will be, employed to provide guidance to governmental actors, the possible actions/choices in the choice set P are referred to as “policies,” with P a particular policy, \( P^* \) a distinct policy, and so forth.

The end product of the SWF framework is a ranking of the choice set P: choice guidance, in the form of this best-to-worst ordering. The methodology arrives at this guidance (for any choice set) by means of three crucial tools: the well-being measure, SWF, and uncertainty module. The well-being measure \( w(\cdot) \) and SWF, together, function to order the outcome set. The uncertainty module, in turn, produces a choice ranking that is appropriately aligned (in consequentialist fashion) with the outcome ranking.

---

16 If an individual lives a certain number of years in one outcome, and a longer or shorter number of years in a second, she exists in both. She “exists” if she is born (comes into being). Her longevity is a determinant of her lifetime well-being (see below, section 2.2.2, for more on lifetime well-being), but doesn’t change the fact of her existence. Thus the fixed-population assumption is fully consistent with using the SWF framework to assess policies that change longevity.
2.2.2 The Well-Being Measure

The well-being measure \( w(\cdot) \) converts each outcome into a “vector” (ordered list) of well-being numbers. This vector has \( N \) entries—one for each member of the population. “\( w_i(x) \)” indicates the well-being number of individual \( i \) in outcome \( x \), as assigned by the well-being measure. A given outcome \( x \) thus corresponds to the vector \((w_1(x), w_2(x), \ldots, w_N(x))\). This vector is a numerical representation of the pattern of well-being in outcome \( x \). Adler (2019, chs. 1-2)

For less mathematical readers, a more concrete illustration may help. Assume that \( N \), the size of the population, is four. So there are four individuals, denoted as 1, 2, 3, and 4. The outcome set \( O = \{x, y, z, \ldots\} \). \( w(\cdot) \) will map each outcome in \( O \) onto a vector with four entries—one for individual 1, one for individual 2, one for individual 3, and one for individual 4. For example, assume that \( w_1(x) = 20 \). The well-being number of individual 1 in outcome \( x \)—as a function of her attribute bundle in \( x \) or her history in \( x \)—is 20. Assume, further, that \( w_2(x) = 60 \). The well-being number of individual 2 in outcome \( x \)—as a function of his attribute bundle in \( x \) or history in \( x \)—is 60. Moreover, \( w_3(x) = 37 \) and \( w_4(x) = 100 \). Then the vector for outcome \( x \) is the four-entry list: \((20, 60, 37, 100)\). There will be a similar four-entry list for each outcome in \( O \).

There is a rich philosophical literature concerning well-being. A range of competing accounts have been proposed and elaborated in this literature. Accounts often fall into one of three categories: hedonic accounts (an individual’s well-being depends on her pains, pleasures, and/or similar mental states that feel good or feel bad); preference accounts (an individual’s well-being depends on the extent to which her preferences are satisfied); and objective-good accounts (well-being depends on the realization of various “objective” goods, “objective” in the sense that whether an individual attains a given good is not merely a matter of what she wants or what she feels). (Adler, 2012, ch. 3; Bykvist, 2016; Fletcher, 2016; Griffin, 1986; Haybron, 2016; Hurka, 2016; Scanlon, 1998, ch. 3; Sumner, 1996).

The \( w(\cdot) \) numbers contain information about well-being comparisons—as per the account of well-being which has been adopted (be it a hedonic, preference-based, or objective-good account). In particular, I’ll assume, those numbers represent intra- and interpersonal comparisons of both well-being levels and differences. \( w_i(x) \geq w_j(y) \) indicates that individual \( i \) in \( x \) is at least as well off as individual \( j \) in \( y \). (If \( i = j \), we have here an intrapersonal comparison of well-being levels. If \( i \neq j \), we have here an interpersonal comparison of well-being levels.) \( w_i(x) - w_j(y) \geq w_k(z) - w_l(zz) \) indicates that the well-being difference between individual \( i \) in \( x \) and individual \( j \) in \( y \) is at least as large as the well-being difference between individual \( k \) in \( z \) and individual \( l \) in \( zz \). (If \( i = j = k = l \), we have here an intrapersonal comparison of well-being differences. If it is not the case that \( i = j = k = l \), we have an interpersonal comparison of well-being differences.)
The SWF framework is consistent with any account of well-being (hedonic, preference-based, objective)—as long as the account doesn’t eschew interpersonal comparisons, and as long as its comparisons can be represented via a well-being measure. Whatever the account, an individual’s well-being will depend on her attributes (or, in the case of a preference account, her history). Thus we can say, generically, that \( w_i(x) \) is a function of individual \( i \)’s attributes/history in \( x \).

See Section 2.5 for further discussion of the different types of well-being comparisons, and for explanation why the SWF framework requires some degree of interpersonal comparability; and see Adler and Decancq, chapter 3, this volume, regarding the construction of an interpersonally comparable well-being measure.

While the SWF framework as presented here is inclusive in the sense of allowing for a wide range of accounts of well-being, it is restrictive in a different way. It assumes that the basis for the ethical assessment of outcomes and policies is individuals’ lifetime well-being. This is implicit in the representation of an outcome as a series of well-being numbers, one number for each person in the population. \( x \) corresponds to \( (w_1(x), w_2(x), \ldots, w_N(x)) \), with \( w_1(x) \) a number measuring the lifetime well-being of individual 1 in \( x \), and so forth. Outcomes are ranked in light of the distributions of lifetime well-being that they embody. If, instead, the basis for ethical assessment were “sublifetime” well-being—well-being during a portion of an individual life, such as a year, a decade, a day, or a moment—then each outcome would correspond to a longer list of sublifetime numbers: for each individual, one number for her sublifetime well-being during each period in that outcome. Although the SWF framework could be specified in this “sublifetime” manner, such an approach is unusual.\(^{17}\)

I have elsewhere presented a lengthy philosophical argument in favor of making lifetime well-being, not sublifetime well-being, the inputs to the SWF—and specifically in favor of lifetime rather than sublifetime prioritarianism. (Adler 2012, ch. 6). In brief: (1) Human persons retain personal identity over their entire lifetimes, by virtue of psychological continuity between sublifetime stages.\(^{18}\) (2) By virtue of such continuity, there is no non-arbitrary way to divide whole lifetimes into sublifetime stages longer than moments. Ordinarily, there is no sharp break in a person’s psychological states at the end of a day, a week, a year, a decade … . Conversely,

\[^{17}\] The sublifetime specification would work as follows. Assume that each individual’s life is divided into sublifetime periods (moments, days, …), and that \( T \) is the maximum number of such periods that anyone lives in any outcome. Let \( w_t^i(x) \) be the well-being of individual \( i \) during period \( t \) of her life, meaning a well-being number if she is alive during that period and a non-numerical symbol if she is not alive. Then \( x \) corresponds to a vector with \( T \times N \) entries: \( (w_{11}^1(x), w_{12}^1(x), \ldots, w_{1T}^1(x), \ldots, w_{N1}^1(x), \ldots, w_{NT}^N(x)) \). Possible rules for ranking \( T \times N \)-entry vectors are specified, and are axiomatized in light of axioms stated in terms of sublifetime numbers (sublifetime Pareto, sublifetime anonymity, sublifetime Pigou-Dalton, and so forth). At least in the theoretical literature on SWFs, such an approach is rarely (if ever) adopted.

\[^{18}\] The case in which a human being has psychological attributes sufficient to be a person, but experiences a psychological break at some point sufficient to break intertemporal personal identity, presents an interesting but exceptional case that can be left aside for purposes of this chapter.
momentary welfarism is quite problematic because (on most accounts of well-being) there are many aspects of well-being that are only realized over longer stretches of time than a moment. (For example, individuals often have preferences for events or experiences that take longer than one moment.). (3) By virtue of intertemporal personal identity, distributional equity should take account of individuals’ lifetime well-being. Imagine that, in outcome $x$, two individuals Avery and Blake have the same sublifetime well-being during period $t$ of their lives. Outcome $y$ is identical to outcome $x$, except that Avery’s well-being in period $t$ has improved by an increment $\Delta w$. Outcome $y^*$ is identical to outcome $x$, except that now it is Blake who is better off by well-being increment $\Delta w$ in period $t$. The sublifetime specification of the SWF framework will be indifferent between $y$ and $y^*$, but it seems quite implausible that the two outcomes are necessarily equally good. For example, imagine that in $x$ Avery is better off than Blake during each period except for $t$. Then it is only fair that Blake receive the benefit; $y^*$ is a better outcome than $y$. The case for $y^*$ being better than $y$ presupposes that Avery remains the same person before, during, and after $t$. (If he and Blake were to experience psychological breaks at the beginning of $t$ and at the end, then $y$ and $y^*$ would be equally good.)

Given space constraints, these arguments will not be recapitulated at greater length here. I take them as persuasive and, throughout the chapter, adopt the whole-lifetime approach—again, that is implicit in mapping an outcome onto one well-being number (a lifetime number) for each person.20

The whole-lifetime approach means that each attribute bundle is a lifetime bundle. It describes a possible combination of characteristics that an individual might have over an entire lifetime. Like the choice of attributes themselves, how fully to describe the intertemporal arc of a life is a matter of modelling tractability.21

2.2.3 The SWF

As already stated, an SWF ($\succeq$) is some rule for ranking well-being vectors. More precisely, it is an “ordering”: a complete, transitive ranking.22

In presenting the various SWFs, it will be useful to have a compact notation for a well-being vector. A lower-case bold letter such as $\mathbf{w}$ or $\mathbf{v}$ will be used to indicate an $N$-entry vector, with $w_i$ the well-being number of individual $i$. (By contrast, the more cumbersome notation

---

19 This assumes that the sublifetime SWF is such as to satisfy the sublifetime version of the Anonymity axiom. Sublifetime Anonymity requires indifference to a rearrangement of sublifetime numbers—which is what has occurred here, since Blake and Avery have simply swapped well-being numbers during period $t$.
20 The reader who prefers the sublifetime approach can translate the formal apparatus of the chapter into sublifetime terms—by substituting $T \times N$-entry vectors for $N$ entry vectors. See note 17 above.
21 See Adler and Decancq, chapter 3, this volume.
22 A quasioordering is a transitive, reflexive, binary relation. The SWF $\succeq$ is assumed to be a complete quasioordering, i.e., an ordering. $W$, the set of well-being vectors ranked by $\succeq$, is either all $N$-dimensional vectors or some orthant of $N$-dimensional space. See Appendix.
\((w_1(x), w_2(x), \ldots, w_N(x))\) is used to indicate the vector associated with a given outcome, in this case \(x\).

\(w \succeq v\) indicates that the SWF ranks vector \(w\) at least as good as vector \(v\). \(w > v\) indicates that the SWF ranks \(w\) strictly better than \(v\). And \(w \sim v\) indicates that the SWF ranks the two vectors as equally good. The \(>\) and \(\sim\) relations are linked to \(\succeq\) as follows: \(w \succeq v\) iff either (a) \(w > v\), or (b) \(w \sim v\). To say that one vector is “at least as good” as a second means that it is \(either\) better than the first \(or\) equally good as the first.

The major SWFs or families thereof are: the utilitarian SWF, prioritarian SWFs, the leximin SWF, rank-weighted SWFs, and sufficientist SWFs.\(^{23}\) The first four are widely used in economic scholarship. The fifth is not, but is included here because it corresponds to an important position in philosophical work on welfare-consequentialism, namely sufficientism. On the different types of SWFs, see Adler (2019, ch. 3); Blackorby, Bossert, and Donaldson (2002; 2005, chs. 2-4); Boadway and Bruce (1984, ch. 5); Bossert and Weymark (2004); d’Aspremont and Gevers (2002); Mongin and d’Aspremont (1998); Weymark (2016).

Rules for all of these SWFs are provided in Table 2.1. The following concrete illustrations may help. Consider the two well-being vectors \(v = (115, 40, 100, 25)\) and \(w = (25, 112, 64, 70)\). The utilitarian SWF assigns each vector a number (“score”) equaling the simple sum of well-being, and compares the vectors according to these scores. \(v\) is assigned the score \(115 + 40 + 100 + 25 = 280\), while \(w\) is assigned the score \(25 + 112 + 64 + 70 = 271\). \(v\) receives the larger sum-of-well-being score (280 > 271), and hence is ranked better than \(w\) by the utilitarian SWF.

\(^{23}\) The leximin SWF is an ordering of well-being vectors, but is not represented by a real-valued function. (In the terminology that will be used below, it is not “score-based.”) This is also true of the sufficientist SWF. The description of these orderings as “social welfare functions” or “SWFs” is thus, strictly speaking, incorrect—but so as to avoid complicating the terminology, I will refer to any \(\succeq\), whether or not score-based, as a “SWF.”
### Table 2.1: The major SWFs

<table>
<thead>
<tr>
<th>SWF Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilitarian SWF</td>
<td>Each vector ( \mathbf{w} ) is assigned a utilitarian score ( S^{Util}(\mathbf{w}) = \mathbf{w}_1 + \ldots + \mathbf{w}_N ). Vectors are ranked in the order of these scores: ( \mathbf{w} \not\succ \mathbf{v} ) if ( S^{Util}(\mathbf{w}) \geq S^{Util}(\mathbf{v}) ).</td>
</tr>
<tr>
<td>Prioritarian SWFs</td>
<td>Each SWF within this family is defined by a strictly increasing and strictly concave function ( g(\cdot) ). Each vector ( \mathbf{w} ) is assigned a prioritarian score ( S^{Prior}(\mathbf{w}) = g(\mathbf{w}_1) + \ldots + g(\mathbf{w}_N) ). Vectors are ranked in the order of these scores: ( \mathbf{w} \not\succ \mathbf{v} ) if ( S^{Prior}(\mathbf{w}) \geq S^{Prior}(\mathbf{v}) ).</td>
</tr>
<tr>
<td>Leximin SWF</td>
<td>Let ( \hat{\mathbf{w}} ) be a vector listing the elements of ( \mathbf{w} ) from smallest to largest. For any two vectors ( \mathbf{w}, \mathbf{v} ): (1) ( \mathbf{w} \sim \mathbf{v} ) if ( \mathbf{w} ) is a permutation of ( \mathbf{v} ); and (2) ( \mathbf{w} \succ \mathbf{v} ) if there is ( j \leq N ) such that ( \hat{\mathbf{w}}_j = \hat{\mathbf{v}}_j ) for all ( i &lt; j ) and ( \hat{\mathbf{w}}_j &gt; \hat{\mathbf{v}}_j ).</td>
</tr>
<tr>
<td>Rank weighted SWFs</td>
<td>Each SWF within this family is defined by a list of ( N ) positive and strictly decreasing weights: ( k_1, k_2, \ldots, k_N ) such that ( k_1 &gt; k_2 &gt; \ldots &gt; k_N &gt; 0 ). Let ( \hat{\mathbf{w}} ) be a vector listing the elements of ( \mathbf{w} ) from smallest to largest. Each vector ( \mathbf{w} ) is assigned a rank-weighted score ( S^{RW}(\mathbf{w}) = k_1 \hat{\mathbf{w}}_1 + \ldots + k_N \hat{\mathbf{w}}_N ). Vectors are ranked in the order of these scores: ( \mathbf{w} \not\succ \mathbf{v} ) if ( S^{RW}(\mathbf{w}) \geq S^{RW}(\mathbf{v}) ).</td>
</tr>
<tr>
<td>Sufficientist SWFs</td>
<td>Each SWF within this family is defined by a threshold level of well-being ( w_{\text{thresh}} ), and some strictly increasing and strictly concave function ( g(\cdot) ). A given well-being vector ( \mathbf{w} ) has two associated vectors: ( \overline{\mathbf{w}} ), the elements of ( \mathbf{w} ) truncated above at ( w_{\text{thresh}} ), and ( \underline{\mathbf{w}} ), the elements of ( \mathbf{w} ) truncated below at ( w_{\text{thresh}} ). For any two vectors ( \mathbf{w}, \mathbf{v} ): (1) ( \mathbf{w} \succ \mathbf{v} ) if either (a) ( g(\overline{\mathbf{w}}_1) + \ldots + g(\overline{\mathbf{w}}_N) &gt; g(\overline{\mathbf{v}}_1) + \ldots + g(\overline{\mathbf{v}}_N) ) or (b) ( g(\underline{\mathbf{w}}_1) + \ldots + g(\underline{\mathbf{w}}_N) = g(\underline{\mathbf{v}}_1) + \ldots + g(\underline{\mathbf{v}}_N) ) and ( \underline{\mathbf{w}}_1 + \ldots + \underline{\mathbf{w}}_N &gt; \underline{\mathbf{v}}_1 + \ldots + \underline{\mathbf{v}}_N ); and (2) ( \mathbf{w} \sim \mathbf{v} ) if ( g(\overline{\mathbf{w}}_1) + \ldots + g(\overline{\mathbf{w}}_N) = g(\overline{\mathbf{v}}_1) + \ldots + g(\overline{\mathbf{v}}_N) ) and ( \underline{\mathbf{w}}_1 + \ldots + \underline{\mathbf{w}}_N = \underline{\mathbf{v}}_1 + \ldots + \underline{\mathbf{v}}_N ).</td>
</tr>
</tbody>
</table>

**Explanation:** The rules for the utilitarian SWF, prioritarian SWFs, and rank-weighted SWFs are stated in a compact form giving necessary and sufficient conditions ("iff" conditions) for it to be the case that \( \mathbf{w} \not\succ \mathbf{v} \). The rules for the lexicmin and sufficientist SWFs could be stated in this compact form. However, so as to facilitate comprehension, I instead state them here as separate necessary and sufficient conditions for \( \mathbf{w} \succ \mathbf{v} \) and for \( \mathbf{w} \sim \mathbf{v} \).

A prioritarian SWF employs some strictly increasing and strictly concave function \( g(\cdot) \) (for short, the prioritarian “transformation” function); assigns each vector a score equaling the sum of transformed well-being numbers; and compares the vectors according to these scores. \( \mathbf{v} \) is assigned the score \( g(115) + g(40) + g(100) + g(25) \), while \( \mathbf{w} \) is assigned the score \( g(25) + g(112) + g(64) + g(70) \). Whether \( \mathbf{v} \) or \( \mathbf{w} \) gets the higher score, in this instance, depends upon which
particular $g(\cdot)$ is chosen as the transformation function. For example, if $g(\cdot)$ is the square root, then $v$ gets the higher score. If $g(\cdot)$ is the logarithm, $w$ does.\footnotemark

A terminological note: in some prior work (Adler 2012, 2019) I have referred to this SWF as a “continuous-prioritarian” SWF. In this chapter and volume, the terminology is simplified. “Prioritarian,” here, means an SWF that sums a strictly increasing and strictly concave function of well-being.

“Prioritarian” SWFs are so-called because they accord greater weight (priority) to well-being changes affecting those at lower well-being levels. Figure 2.1 illustrates why the use of a strictly increasing and strictly concave transformation function means that a prioritarian SWF will indeed do this. Let $w_L$ be the well-being level of a worse-off person, and $w_H$ the well-being level of a better-off person: $w_L < w_H$. Imagine changing the worse-off person’s well-being by adding a given amount of well-being, $\Delta w$, or instead changing the better-off person’s well-being by adding the same amount. Because $g(\cdot)$ is strictly increasing, each change increases the prioritarian score. Because $g(\cdot)$ is strictly concave, the first change produces a larger increase in the prioritarian score than the second.

\footnotetext{24 The straight prioritarian formula assigns each well-being vector $w$ the score $\sum_{i=1}^{N} g(w_i)$. Sometimes a different prioritarian formula is used, the “equally distributed equivalent” (EDE) formula: $g^{-1}\left(\frac{\sum_{i=1}^{N} g(w_i)}{N}\right)$. The two formulas are ordinally equivalent: they rank well-being vectors the same way. The choice between the straight and EDE formulas becomes significant in other contexts: for purposes of the uncertainty module, and for purposes of measuring social welfare on a ratio or translation scale. See Appendix, 2.A.2; Decancq and Schokkaert, chapter 5, this volume.}
The leximin SWF ranks vectors according to the lowest well-being levels in each; if these are equal, the second-lowest; and so forth. The lowest well-being level in \( v \) (25, the well-being number of the fourth person) is the same as the lowest well-being level in \( w \) (25, the well-being number of the first person). However, the second-lowest well-being level in \( v \) (40, the well-being of the second person) is below the second-lowest well-being level in \( w \) (64, the well-being
of the third person), and so the leximin SWF prefers \( w \). Leximin is related to the notion of “maximin” introduced by John Rawls in *A Theory of Justice*.\(^{25}\) (Rawls 1999, first published in 1971).

A rank-weighted SWF uses a list of \( N \) fixed and positive weights, \( k_1, k_2, \ldots, k_N \), which are strictly decreasing: \( k_1 > k_2 > \ldots > k_N > 0 \). Each vector is assigned a score by multiplying the lowest well-being number in that vector times \( k_1 \) plus the second-lowest number times \( k_2 \) plus \ldots plus the highest number times \( k_N \). For example, vector \( v \) would be assigned the score \( k_1 \times 25 + k_2 \times 40 + k_3 \times 100 + k_4 \times 115 \), while \( w \) would be assigned the score \( k_1 \times 25 + k_2 \times 64 + k_3 \times 70 + k_4 \times 112 \). In this instance, which vector is ranked higher depends upon the weights \( k_1, k_2, k_3 \) and \( k_4 \).

In a number of tables below, I illustrate the properties of rank-weighted SWFs using the rank-weighted SWF with “integer weights,” by which I mean that: \( k_1 = N, k_2 = N - 1, \ldots, k_N = 1 \).

A sufficientist SWF is defined by specifying some well-being threshold \( w^{\text{thresh}} \) and a strictly increasing and strictly concave transformation function \( g(\cdot) \). The SWF is prioritarian below the threshold; gives absolutely priority to those below the threshold, over those above; and is utilitarian above the threshold. This is accomplished via a two-step rule: compare vectors by applying the prioritarian rule to well-being entries truncated above at \( w^{\text{thresh}} \); if the vectors are ranked equal at the first step, compare them by applying the utilitarian rule to well-being entries truncated below at \( w^{\text{thresh}} \). For example, let \( w^{\text{thresh}} = 110 \) and let \( g(\cdot) \) be some strictly increasing and strictly concave function. The vectors \( v \) and \( w \) above are compared as follows. Let \( \overline{v} \) and \( \overline{w} \) denote the entries in those vectors truncated above at \( w^{\text{thresh}} \), and \( v \) and \( w \) their entries truncated below at \( w^{\text{thresh}} \). So \( \overline{v} = (110, 40, 100, 25) \), \( \overline{w} = (25, 110, 64, 70) \), \( v = (115, 110, 110, 110) \), \( w = (110, 112, 110, 110) \). Then we start by comparing \( g(110) + g(40) + g(100) + g(25) \) to \( g(25) + g(110) + g(64) + g(70) \). Depending on the \( g(\cdot) \) function, the first sum might be greater (if so, the sufficientist SWF ranks \( v \) over \( w \) ), the second might be greater (if so, it ranks \( w \) over \( v \)).

---

\(^{25}\) Rawls was not a welfare-consequentialist. Maximin as articulated in *A Theory of Justice* in the form of the “Difference Principle” applies to the distribution of “primary goods,” not well-being: it says to maximize the bundle of primary goods held by the worst off. Maximin, transposed to the SWF format, takes the following form: \( w \) is better than \( v \) iff the lowest well-being level in \( w \) is greater than the lowest well-being level in \( v \).

As a matter of theory, maximin is not a plausible SWF because it violates the Strong Pareto axiom (see Section 2.3). If the lowest well-being level in \( w \) is equal to the lowest well-being level in \( v \), but some individuals are better off in \( w \) than \( v \) and none are worse off, then Strong Pareto requires \( w \) to be ranked above \( v \)—but maximin will be indifferent. Like maximin, leximin is such that an improvement in the well-being of the worst off is sufficient for an ethical improvement; unlike maximin, leximin satisfies Strong Pareto. For example, both maximin and leximin prefer \( w^* = (4, 25, 80, 60) \) to \( v^* = (3, 100, 100, 100) \) since the worst off improves from 3 to 4. However, maximin is indifferent between \( v = (15, 3, 9, 11) \) and \( w = (15, 3, 10, 14) \), since the worst off remains at 3; Strong Pareto applies and requires a preference for \( w \); and leximin indeed prefers \( w \).

Despite the axiomatic difficulties of the maximin SWF, it is sometimes used in applications. This can be justified by seeing maximin as a tractable approximation to leximin. Leximin is not score-based (see below, this subsection), while maximin is score-based—the score assigned to a vector is just the well-being number of the worst off individual—which makes it more tractable.
or they might be equal. In this last case we move on to the second step, comparing \( 115 + 110 + 110 + 110 = 110 + 112 + 110 + 110 \)—so that the SWF ranks \( v \) over \( w \).

Note that some but not all SWFs are “score-based.” These SWFs employ some mathematical formula for assigning each vector a numerical score, and rank the vectors according to these scores. The utilitarian SWF, prioritarian SWFs, and rank-weighed SWFs are score-based. By contrast, the leximin SWF and sufficientist SWFs are not score-based. This difference, and the underlying axiomatics, will be further pursued below.

The SWF framework orders the outcome set via the combination of \( w(\cdot) \) and \( \succeq \). Let’s call this the “master recipe.”

**The Master Recipe**: \( x \) at least as good as \( y \) iff \( (w_1(x), \ldots, w_N(x)) \succeq (w_1(y), \ldots, w_N(y)) \)

Outcomes are ordered according to the SWF’s ranking of their corresponding well-being vectors. Note how this implements welfare-consequentialism: the ethical comparison of any two outcomes hinges on a comparison of the patterns of well-being in the two (as represented numerically via well-being vectors). Note also that the Master Recipe is fully generic. How in fact a given outcome set will be ordered hinges on which well-being measure is used, and on which SWF is used.

### 2.2.4 The Uncertainty Module

Let’s discuss, finally, the SWF’s “uncertainty” module. This tells us how to move from the ranking of outcomes according to the SWF (in conjunction with a given well-being measure) to the ranking of the choice set. On SWFs under uncertainty, see Adler (2012, ch. 7; 2019, chs. 3-4); Fleurbaey (2010); Mongin and Pivato (2016).

A standard device in decision theory is to represent choice under uncertainty via the state-of-nature set up. (Gilboa, 2009, chs. 10-12; Joyce and Gibbard, 1998). There’s a group of possible “states of nature.” One or another of these is the actual state; the decisionmaker doesn’t know which. The states should be such as to be causally independent of the choice at hand.\(^{26}\) Moreover, they should be sufficient, in combination with the decisionmaker’s choice, to yield one or another outcome. That is, each choice in the choice set, in conjunction with a given state, produces some outcome. This is the outcome that would occur, were that choice to be selected

\(^{26}\)That is: the decisionmaker’s selection from the choice set doesn’t change which state is actual. Specifically, then, a possible state of nature should be thought of as a combination of (a) possible facts about the world at or prior to the time of choice and (b) possible causal laws. The “actual” state is, thus, a combination of possible facts about the world at or prior to the time of choice, and possible causal laws, which are such as to be true facts about the decisionmaker’s world (facts about the world in which she actually finds herself) and true causal laws (the causal laws in her actual world).
and that state to be the actual state. The decisionmaker’s uncertainty regarding which state is actual is expressed probabilistically: the states are assigned probabilities, summing to 1.\textsuperscript{27}

Let’s now integrate this state-of-nature set-up into the SWF framework. As illustrated in Table 2.2, a given policy in the choice set $P$ yields some outcome in each state. Thus, for example, policy $P$ produces outcome $x$ in state $s$; outcome $y$ in state $s^+$; $z$ in state $s^{++}$; and $zz$ in state $s^{+++}$. We might say that each policy is a state-by-state array of outcomes. Moreover, with the well-being measure $w(\cdot)$ in hand, each policy becomes a state-by-state array of well-being vectors. $P$ in state $s$ produces some outcome $x$; $x$ is converted by the well-being measure into a well-being vector; and hence $P$ is assigned that well-being vector in state $s$.

Table 2.2: The state-of-nature format for representing uncertainty

<table>
<thead>
<tr>
<th>States of nature</th>
<th>$s$</th>
<th>$s^+$</th>
<th>$s^{++}$</th>
<th>$s^{+++}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilities</td>
<td>$\pi(s)$</td>
<td>$\pi(s^+)$</td>
<td>$\pi(s^{++})$</td>
<td>$\pi(s^{+++})$</td>
</tr>
<tr>
<td>Policy $P$</td>
<td>$x$</td>
<td>$y$</td>
<td>$z$</td>
<td>$zz$</td>
</tr>
<tr>
<td>Policy $P^*$</td>
<td>$x$</td>
<td>$z$</td>
<td>$zz$</td>
<td>$z$</td>
</tr>
<tr>
<td>Policy $P^{**}$</td>
<td>$z$</td>
<td>$y$</td>
<td>$y$</td>
<td>$x$</td>
</tr>
</tbody>
</table>

Explanation: In this example, there are four states of nature: $s$, $s^+$, $s^{++}$, $s^{+++}$. The probability of a given state is denoted with “$\pi(\cdot)$”—for example, the probability of state $s^+$ is $\pi(s^+)$. The cumulative probability of the states is 1. Thus, here, $\pi(s) + \pi(s^+) + \pi(s^{++}) + \pi(s^{+++}) = 1$. In this example, there are four possible outcomes ($x$, $y$, $z$, and $zz$). A given policy associates each state with one of the outcomes; this is the outcome that would occur, were the policy to be chosen and that state to be the actual state of nature. For example, policy $P^{**}$ produces outcome $z$ in state $s$, outcome $y$ in state $s^+$, outcome $y$ in state $s^{++}$, and outcome $x$ in state $s^{+++}$.

The uncertainty module for a given SWF is a procedure for ranking policies (each seen as a state-by-state array of well-being vectors). Interestingly, each SWF has multiple uncertainty modules. Of course, and trivially, every module for a given SWF should converge in its ranking of “degenerate” policies—a “degenerate” policy being one that yields the same vector in each state. (If all policies in the choice set are degenerate, there is no genuine, i.e., ethically relevant, uncertainty: no uncertainty regarding what the well-being pattern will be, were a particular policy to be selected.) But this basic requirement doesn’t resolve how the SWF should be used to rank policies when there is genuine uncertainty.

The fact that a particular SWF has multiple uncertainty modules will be illustrated below in Section 2.7, where I describe three distinct modules for the prioritarian SWF.\textsuperscript{28}

\textsuperscript{27} Sometimes, the term “uncertainty” is used to indicate the absence of precise probabilities, by contrast with “risk.” In this chapter, however, “uncertainty” simply means the decisionmaker’s absence of certainty regarding which outcome would result from a given policy; and I assume, indeed—in line with classical decision theory—that the decisionmaker can always express her absence of certainty by an assignment of precise probabilities to states. How to apply an SWF absent such probabilities is beyond the scope of the chapter.
2.2.5 A Note on the Integral Representation

The theoretical literature on SWFs generally employs the formalism set forth here. There are a finite number of individuals in each outcome; an outcome corresponds to a well-being vector with a finite number of entries, one for each person. For short, call this the “vector” version of the SWF framework.

A different version of the framework employs an integral representation of the pattern of well-being. There are an uncountable infinity of individuals. An outcome corresponds to a well-being distribution across the uncountable population. Each such distribution is specified by a density function $f(\cdot)$. For a distribution with density $f(\cdot)$, the proportion of the population with well-being levels between $w^*$ and $w^{**}$ is
\[
\int_{w^*}^{w^{**}} f(w) \, dw.
\]
A SWF is a rule for ranking distributions.

The utilitarian SWF ranks distributions according to the formula
\[
\int_{-\infty}^{\infty} w f(w) \, dw.
\]
The prioritarian SWF uses the formula
\[
\int_{-\infty}^{\infty} g(w) f(w) \, dw.
\]
The integral formulation is the cornerstone of the SWF-based literature on optimal income taxation. (Salanie 2011; Tuomala 2016). Indeed, the optimal taxation chapter in this volume (Tuomala and Weinzierl, chapter 4) uses this formulation. It also figures elsewhere in the volume. (Decancq and Schokkaert, chapter 5; Ooghe, chapter 9).

The vector version has the theoretical virtue of mapping directly onto the structure of individuals and well-being. Individuals are discrete beings, each with a well-being level. Each possible world contains a finite number (or if not a finite number, then at most a countable infinity) of these discrete beings, not an uncountable infinity. Further, it is more flexible theoretically than the integral approach, in that certain axiomatic possibilities are ruled out by the

---

28 It’s worth noting that the decisionmaker implementing the SWF framework doesn’t literally need to specify states and a mapping from states to outcomes for each policy. Instead, she might circumvent the explicit representation of states and, instead for each policy, simply assign a probability to each outcome given the choice of that policy. However, such assignment should be consistent with the state framework. These should be causal probabilities. The probability of $x$ given $P$ should be the probability of $P$ causing $x$ to occur. That is, this probability should be the cumulative probability of the various possible states, each state causally independent of $P$, which are such that—were one of these states to obtain—the choice of $P$ would cause $x$ to occur.

29 This chapter assumes throughout a finite population of individuals in each outcome. This is true both for the main case of a fixed and finite population ($N$) in all outcomes; and for the extension, in Section 2.11, to allow for variable population (whereby it is allowed that the number of individuals each of whom exists in at least one outcome may be countably infinite, but the number in each outcome is finite). There is, in addition, an SWF-based literature on “infinite population” that allows for a countable infinity of individuals in an outcome—for example, a countable infinity of future generations. See Adler (2009, pp. 1511-19), reviewing this literature.
latter.\textsuperscript{30} On the other hand, in some policy contexts, it may be easier to work with integrals than with finite sums—this seems to have been the experience, for example, of optimal tax scholars.

Although this chapter uses the vector formulation, the reader should not be surprised when he or she sees the integral formulation elsewhere in this volume and in the literature; and much of the analysis here could be transposed to that formulation.

2.2.6 A Note on Generalized-Lorenz Dominance

One way to employ the SWF framework for policy analysis is to choose a particular well-being measure $w(\cdot)$, a particular SWF, and a particular uncertainty module for that SWF, and then apply this specific measure-SWF-module combination in the policy domain of interest. This approach has the advantage of yielding a complete ranking of outcomes and policies. However, it requires a series of difficult ethical judgments—namely those involved in ruling out all other well-being measures, SWFs, and modules.

A different, less demanding tactic is to leave the choice of well-being measure, SWF, and/or module partly unspecified. We could follow this route in an ad hoc manner. For example, pick a well-being measure, a family of SWFs, and a particular type of module for each SWF. See which policy comparisons are robust to the choice of any SWF within the family, and which policy comparisons are sensitive to that choice.

It would be useful, however, to have theoretical (not merely ad hoc) results regarding the application of the SWF framework without a fully specified well-being measure, SWF, and module—and indeed there is a theory literature on this topic. The literature goes under the heading of “dominance.” It shows that if well-being vector $w$ “dominates” vector $v$ in some specified sense, then $w$ is preferred to $v$ by every SWF within some class, and vice versa.\textsuperscript{31} (Lambert 2001, chs. 3-4; Cowell 2016; Chakravarty 2009, ch. 1; Shorrocks 1983).

A headline finding of the “dominance” literature concerns “generalized-Lorenz” (GL) dominance. Consider any well-being vector $w$ of length $N$. The generalized-Lorenz curve maps 1 onto the lowest well-being number in $w$, 2 onto the sum of the two lowest entries in $w$, 3 onto the sum of the three lowest entries, …, $N$ onto the sum of the $N$ lowest entries (total well-being).\textsuperscript{32} Vector $w$ GL-dominates vector $v$ if the generalized-Lorenz curve for $w$ is sometimes

\textsuperscript{30} For example, Anonymity flows automatically from the integral approach. Well-being distributions are fully characterized by the density function, and so two distributions with the same density are ranked equally good regardless of which individuals are at the different percentiles of the distributions; while the vector approach allows for a violation of Anonymity. See (Cowell 2000, pp. 91-92), discussing the two approaches.

\textsuperscript{31} The dominance literature is often formulated in terms of income vectors and social evaluation functions applied to income vectors; but dominance results of this sort can easily be transposed to vectors of well-being numbers and SWFs.

\textsuperscript{32} We are assuming, as throughout this chapter until Section 2.11, that all vectors being compared have the same length $N$. 
above the curve for \( v \), and never below. A theorem of Shorrocks (1983) shows that \( w \) GL-
ominates \( v \) iff every prioritarian SWF ranks \( w \) above \( v \).

Ranking well-being vectors by the relation of GL-dominance has the advantage of
avoiding contestable questions regarding the choice of prioritarian transformation function \( g(\cdot) \).
If one vector GL-dominates a second, then we know that every prioritarian—regardless of the
functional form of her preferred transformation function and its degree of concavity—will count
the first vector as better. A disadvantage of this approach is that it can be quite “gappy.” Note
that the relation of GL-dominance is a partial, rather than complete ordering. It can be the case
that the generalized-Lorenz curves of vectors \( w \) and \( v \) cross, so that neither GL-dominates the
other. (Indeed, by the Shorrocks theorem, this will occur whenever one prioritarian SWF prefers
\( w \) and a second \( v \).)

GL-dominance is further discussed in Brunori, Ferreira, and Peragine, chapter 11, this
volume, and is the basis for the empirical component of that chapter.

2.3 Axiomatic Characterization of SWFs

There are many different SWFs (indeed, an infinity!). We can structure ethical debate
about the appropriate SWF by means of axioms. An axiom constrains the ranking of well-being
vectors. By endorsing an axiom, we rule out SWFs inconsistent with it. On the axiomatic
characterization of SWFs, see Adler (2019, ch. 3); Blackorby, Bossert, and Donaldson (2002;
2005, chs. 2-4); Boadway and Bruce (1984, ch. 5); Bossert and Weymark (2004); d’Aspremont
and Gevers (2002); Mongin and d’Aspremont (1998); Weymark (2016).

Two fundamental axioms, widely accepted by scholars who work with SWFs, and
satisfied by all of the SWFs mentioned earlier (utilitarian, prioritarian, leximin, rank-weighted,
and sufficientist) are: the Pareto axiom (which in turn has two parts, Pareto Indifference and
Strong Pareto); and the Anonymity axiom.

I generally provide informal statements of the axioms; formal statements are provided in
the Appendix. I also illustrate each axiom with an example.

**Pareto.** (1) **Pareto Indifference.** If each person is equally well off in vector \( w \) as in vector
\( v \), then \( w \sim v \). Example: \((12, 25, 80, 3) \sim (12, 25, 80, 3)\)

(2) **Strong Pareto.** If each person is at least as well off in \( w \) as in \( v \), and at least
one person is strictly better off, then \( w > v \). Example: \((3, 4, 10, 13) > (3, 4, 10, 12)\)

**Anonymity.** If the well-being numbers in \( w \) are a permutation (rearrangement) of the
well-being numbers in \( v \), then \( w \sim v \). Example: \((7, 12, 4, 60) \sim (12, 60, 4, 7)\)
Pareto Indifference, as mentioned earlier, is the defining feature of welfare consequentialism. All SWFs satisfy it automatically. The other part of the Pareto axiom, Strong Pareto, is not automatically satisfied. But it should seem very plausible to a welfare consequentialist. The fact that some individuals have a higher well-being level in outcome $x$ than $y$ is, prima facie, a reason to count the first outcome better. If none have a lower well-being level, what countervailing consideration would justify the conclusion that the first outcome is not better? Perhaps there is some non-welfare fact about the two outcomes such that, on balance, $y$ is better than $x$ or equally good. Non-welfarist consequentialists might accept this line of argument, but welfare-consequentialists can’t—since they deny that anything other than the pattern of welfare drives the outcome ranking.

Anonymity captures the ethical attitude of impartiality as between all members of the population. To say that $w$ is a permutation (rearrangement) of the well-being numbers in $v$ is just to say they contain the very same pattern of well-being—in the sense that, for any well-being level, the number of individuals in $w$ at that level equals the number in $v$. The two vectors differ only as regards which particular individual is at a given well-being level. But, from an impartial point of view, that difference shouldn’t matter.

Three further axioms are more controversial: the axioms of Pigou-Dalton (an equity axiom), Separability, and Continuity. These clarify what prioritarianism has in common with other SWFs, and what makes it distinctive. A prioritarian SWF satisfies all three of these axioms. None of the other SWFs that have been mentioned do. This is no coincidence: it can be shown that if an SWF satisfies Pigou-Dalton, Separability, and Continuity, along with the uncontroversial axioms of Pareto and Anonymity, then it must be prioritarian.$^{33}$

Pigou-Dalton.$^{34}$ Assume that $i$ is better off than $j$ in vector $v$; that in $w$ $i$’s well-being has decreased by $\Delta w$, while $j$’s well-being has increased by the very same amount, and the gap between the two individuals has decreased; and that every other person is equally well off in $w$ as she is in $v$. Then $w \succ v$. Example: $(3, 6, 8, 12) \succ (3, 4, 10, 12)$

Separability. Assume that some individuals are “affected” with respect to the $w/v$ comparison (each such individual is either better off in $w$ or better off in $v$), while others

$^{33}$ See note 37 below.

$^{34}$ The Pigou-Dalton principle here is stated in terms of a transfer of well-being from someone with more well-being to someone with less. This is the appropriate form of the principle for a constraint on the ranking of well-being vectors. This version of the principle should be distinguished from a Pigou-Dalton principle framed in terms of some non-well-being attribute (an input to well-being)—for example, the Pigou-Dalton principle stated in terms of a transfer of income from someone with more income to someone with less, which is a mainstay of the literature on income inequality. (Cowell, 2016).

The Pigou-Dalton principle with respect to well-being might be stated, as here, in terms of gap-diminishing transfers; or rather as a principle limited to non-rank-switching transfers. (Adler, 2012 ch. 5). The version here implies the latter version; and, assuming Anonymity, the latter implies the version here. Thus, assuming Anonymity, the two are equivalent.
are “unaffected” (each such individual is equally well off in \( w \) as she is in \( v \)). Let two other vectors, \( w^* \) and \( v^* \), be such that: (1) each individual who is affected in the \( w/v \) comparison is at the same well-being level in \( w^* \) as in \( w \), and the same well-being level in \( v^* \) as in \( v \); and (2) each individual who is unaffected in the \( w/v \) comparison is also unaffected in the \( w^*/v^* \) comparison. Then the \( w/v \) ranking is the same as the \( w^*/v^* \) ranking, i.e., \( w \succ v \) iff \( w^* \succ v^* \).

**Example:** \((7, 100, 23, 7) \succ (4, 100, 23, 12) \iff (7, 7, 7, 7) \succ (4, 7, 7, 12)\)

**Continuity.** If \( w \succ v \), then there is some region around \( w \) (perhaps small) such that for every \( w^* \) in this region, \( w^* \succ v \). Similarly, if \( v \succ w \), then there is some region around \( w \) (perhaps small) such that for every \( w^* \) in this region, \( v \succ w^* \). **Example:** If \((1, 3, 100, 100) \succ (1, 3, 6, 8)\), then \((1, 3 \pm \varepsilon, 100, 100) \succ (1, 3, 6, 8)\) for \( \varepsilon \) sufficiently small.

Utilitarianism fails Pigou-Dalton, but satisfies Separability and Continuity. Leximin, like prioritarianism, satisfies Pigou-Dalton and Separability, but it fails Continuity. Rank-weighted SWFs satisfy Pigou-Dalton and Continuity, but fail Separability. Sufficientist SWFs don’t fully satisfy Pigou-Dalton,\(^\text{35}\) satisfy Separability, but fail Continuity.\(^\text{36}\)

Tables 2.3, 2.4, and 2.5 illustrate the properties of the various SWFs with respect to these three axioms.

Note now that the utilitarian SWF and prioritarian SWFs are axiomatically similar up to Pigou-Dalton, the equity axiom. They both satisfy Pareto, Anonymity, Separability, and Continuity; the axiomatic difference (an important one, to be sure), is that prioritarian SWFs also satisfy Pigou-Dalton.

In fact, by combining Pareto, Anonymity, Separability, and Continuity, we characterize a class of SWFs: generalized-utilitarian SWFs. Generalized-utilitarian SWFs are score-based, and assign each vector \( w \) a score equaling \( f(w_1) + f(w_2) + \ldots + f(w_N) \), with \( f(\cdot) \) a strictly increasing and continuous function. Generalized-utilitarian SWFs clearly satisfy Pareto, Anonymity, Separability, and Continuity, and the converse can also be demonstrated: if an SWF satisfies these four axioms, then it must take the generalized-utilitarian form. (Adler 2018, appendix; Blackorby, Bossert and Donaldson 2005, p. 116; Bossert and Weymark 2004, p. 1159). The utilitarian SWF and prioritarian SWFs are members of the generalized-utilitarian class (there are others as well). In the case of utilitarianism, the \( f(\cdot) \) function is the identity function: \( f(w_i) = w_i \).

---

\(^{35}\) They satisfy it only if the worse-off individual is below the threshold. By contrast, if \( w \) is reached from \( v \) via a gap-diminishing pure transfer from a better-off individual above the threshold to a worse-off individual at or above the threshold, leaving everyone else unaffected, a sufficientist SWF will count \( w \) and \( v \) as equally good.

\(^{36}\) This chapter uses the Pigou-Dalton axiom as the equity axiom for SWFs. Two other equity concepts also widely used by welfare economists are quasi-concavity and S-concavity. (Chakravarty 2009, ch. 1; Sen 1973; Simon and Blume 1994, pp. 522-27). A discussion of these concepts and how they relate to the Pigou-Dalton axiom will not be undertaken here. I will simply note that prioritarian SWFs are also equity regarding in these senses: any prioritarian SWF is both strictly quasi-concave and strictly S-concave.
In the case of prioritarianism, the $f(\cdot)$ function is the prioritarian transformation function, i.e., $f(\cdot) = g(\cdot)$, with $g(\cdot)$ strictly increasing and strictly concave.\textsuperscript{37}

**Table 2.3: The Pigou-Dalton axiom**

<table>
<thead>
<tr>
<th></th>
<th>Vector w</th>
<th>Vector v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ariel</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Ben</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Camilla</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>Diego</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Emile</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Utilitarian score</th>
<th>Prioritarian score (√)</th>
<th>Rank weighted score (integer weights)</th>
<th>Leximin</th>
<th>Sufficientist:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>181</td>
<td>27</td>
<td>331</td>
<td>v &gt; w</td>
<td>$w^{\text{thresh}} = 50$ and $\sqrt{\cdot}$ v &gt; w</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.47</td>
<td></td>
<td></td>
<td>$w^{\text{thresh}} = 15$ and $\sqrt{\cdot}$ v &gt; w</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$w^{\text{thresh}} = 2$ and $\sqrt{\cdot}$ v ~ w</td>
</tr>
</tbody>
</table>

**Explanation:** The Pigou-Dalton axiom requires that $v$ be ranked above $w$. The table illustrates the ranking of these vectors by the utilitarian SWF; a prioritarian SWF with a square root transformation function; a rank-weighted SWF with integer weights; leximin; and a sufficientist SWF with a square root transformation function and various thresholds.

For the score-based SWFs (utilitarian, prioritarian, rank-weighted), the table displays the scores assigned to $v$ and $w$; the ranking of the vectors is in the order of these scores. For the leximin and sufficientist SWFs, which are not score-based, the table directly displays the ranking of the two vectors. A similar approach is followed in tables 2.4 and 2.5 below.

---

\textsuperscript{37} The prioritarian transformation function $g(\cdot)$ is also continuous. If the domain of $g(\cdot)$ is an open interval, this is automatic, since concave functions are continuous on open intervals. Because $g(\cdot)$ is strictly increasing, continuous, and strictly concave, it is straightforward that the prioritarian SWF satisfies Pareto, Anonymity, Separability, Continuity, and Pigou-Dalton. The converse is also true. The first four axioms ensure that the SWF takes the generalized-utilitarian form—summing well-being transformed by a strictly increasing and continuous function $f(\cdot)$. Adding Pigou-Dalton entails that $f(\cdot)$ is strictly concave. (Adler 2019, p. 273).
### Table 2.4: The Separability axiom

<table>
<thead>
<tr>
<th></th>
<th>Vector ( w )</th>
<th>Vector ( v )</th>
<th>Vector ( w^* )</th>
<th>Vector ( v^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ariel</td>
<td>30</td>
<td>32</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>Ben</td>
<td>60</td>
<td>57</td>
<td>60</td>
<td>57</td>
</tr>
<tr>
<td>Camilla</td>
<td>100</td>
<td>100</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>Diego</td>
<td>101</td>
<td>101</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>Emile</td>
<td>102</td>
<td>102</td>
<td>43</td>
<td>43</td>
</tr>
</tbody>
</table>

Utilitarian score 393 392 216 215
Prioritarian score (√) 43.37 43.36 32.66 32.65
Rank weighted score (integer weights) 994 992 586 593
Leximin \( v \succ w \) \( v^* \succ w^* \)
Sufficientist: \( w^{\text{thresh}} = 50 \) and \( \sqrt{v} \succ w \) \( v^* \succ w^* \)

**Explanation:** The Separability axiom requires that \( w \succeq v \) iff \( w^* \succeq v^* \). The table illustrates that the rank-weighted SWF fails Separability, while the other SWFs satisfy it. In this table, as in Table 2.4, the square root function is used as the transformation function for the prioritarian and sufficient SWFs; and the weights for the rank-weighted SWF are integer weights.

### Table 2.5: The Continuity Axiom

<table>
<thead>
<tr>
<th></th>
<th>Vector ( w )</th>
<th>Vector ( v )</th>
<th>Vector ( w^* )</th>
<th>Vector ( v^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ariel</td>
<td>4</td>
<td>4</td>
<td>4.1</td>
<td>4</td>
</tr>
<tr>
<td>Ben</td>
<td>9</td>
<td>16</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Camilla</td>
<td>81</td>
<td>25</td>
<td>81</td>
<td>25</td>
</tr>
<tr>
<td>Diego</td>
<td>121</td>
<td>100</td>
<td>121</td>
<td>100</td>
</tr>
<tr>
<td>Emile</td>
<td>144</td>
<td>225</td>
<td>144</td>
<td>225</td>
</tr>
</tbody>
</table>

Utilitarian score 359 370 359.1 370
Prioritarian score (√) 37 36 37.02 36
Rank weighted score (integer weights) 685 584 685.5 584
Leximin \( v \succ w \) \( w^* \succ v \)
Sufficientist: \( w^{\text{thresh}} = 7 \) and \( \sqrt{v} \succ w \) \( w^* \succ v \)

**Explanation:** Continuity requires that if \( w \succ v \) then this is also true for every vector sufficiently close to \( w \), and that if \( v \succ w \) then this is also true for every vector sufficiently close to \( w \). The table illustrates that Continuity is satisfied by the utilitarian SWF, prioritarian SWFs, and rank-weighted SWFs, but not the leximin SWF or sufficientist SWFs. In this table, as in Table 2.3 and 2.4, the square root function is used as the transformation function for the prioritarian and sufficient SWFs; and the weights for the rank-weighted SWF are integer weights.
2.4 Defending Generalized-Utilitarian SWFs

Generalized-utilitarian SWFs are considerably more tractable than other types of SWFs. Let’s contrast them, first, with SWFs that fail Separability (such as rank-weighted SWFs); and, second, with SWFs that satisfy Separability but fail Continuity (such as the leximin SWF and sufficientist SWFs).

Separability is a precondition for an axiom regarding the uncertainty module, which I’ll term “Policy Separability.” (Adler 2019, pp. 285-88). Only SWFs that satisfy Separability can have an uncertainty module that satisfies Policy Separability. If an SWF fails Separability, then its uncertainty module necessarily fails Policy Separability.³⁸

Policy Separability. To say that one or more individuals are “sure to be unaffected” as between two policies \( P \) and \( P^* \) means: each such individual with \( P \) has the same well-being level in any given state that she does with \( P^* \). An uncertainty module satisfies the axiom of Policy Separability if its ranking of any two policies, \( P \) and \( P^* \), is invariant to the state-by-state well-being levels of individuals who are sure to be unaffected as between \( P \) and \( P^* \).

For an illustration of Policy Separability, see Table 2.6.

---

³⁸ This is straightforward. A given well-being vector \( w \) is equivalent to a “degenerate” policy: one that yields vector \( w \) in every state. If an SWF fails Separability, then any uncertainty module for that SWF necessarily fails Policy Separability (at least) in ranking the set of all degenerate policies, or any set that includes all degenerate policies.
Table 2.6: Policy Separability

<table>
<thead>
<tr>
<th></th>
<th>Policy $P$</th>
<th></th>
<th>Policy $P^*$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>State $s$</td>
<td>State $s'$</td>
<td>State $s$</td>
<td>State $s'$</td>
</tr>
<tr>
<td></td>
<td>$\pi(s)=.5$</td>
<td>$\pi(s')=5$</td>
<td>$\pi(s)=.5$</td>
<td>$\pi(s')=5$</td>
</tr>
<tr>
<td>Charles</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Deb</td>
<td>10</td>
<td>12</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Ernie</td>
<td>12</td>
<td>100</td>
<td>12</td>
<td>100</td>
</tr>
<tr>
<td>Utilitarian score</td>
<td>23</td>
<td>114</td>
<td>68.5</td>
<td>Expected</td>
</tr>
<tr>
<td>Prioritarian score (√)</td>
<td>7.63</td>
<td>14.88</td>
<td>11.25</td>
<td>21</td>
</tr>
<tr>
<td>EDE (√)</td>
<td>6.46</td>
<td>24.60</td>
<td>15.53</td>
<td>6.59</td>
</tr>
<tr>
<td>Rank weighted score (integer weights)</td>
<td>35</td>
<td>130</td>
<td>82.5</td>
<td>34</td>
</tr>
</tbody>
</table>

|                | Policy $P^*$ |          | Policy $P^{**}$ |          |
|                | State $s$  | State $s'$ | State $s$  | State $s'$ |
|                | $\pi(s)=.5$ | $\pi(s')=5$ | $\pi(s)=.5$ | $\pi(s')=5$ |
| Charles        | 1           | 2         | 4           | 4         |
| Deb            | 10          | 12        | 5           | 8         |
| Ernie          | 5           | 6         | 5           | 6         |
| Utilitarian score | 16        | 20       | 18          | Expected  |
| Prioritarian score (√) | 6.40 | 7.33 | 6.86 | 14 | 18 | 16 |
| EDE (√)       | 4.55        | 5.97      | 5.26        | 6.47      | 7.28      | 6.88        |
| Rank weighed score (integer weights) | 23 | 30 | 26.5 | 27 | 32 | 29.5 |

Explanation: Ernie is sure to be unaffected as between $P$ and $P^*$: he has the same well-being level in state $s$ with each policy, and the same well-being level in state $s'$ with each policy. Policy Separability requires that the $P/P^*$ ranking be invariant to these state-by-state levels. For example, it requires that the $P/P^*$ ranking be the same as the $P^*/P^{**}$ ranking.

The table is meant to illustrate section 2.7 of this chapter as well as the current section. It does so by showing how various uncertainty modules, described in Section 2.7 of this chapter, fare with respect to Policy Separability. The axiom is satisfied by UUU, the standard module for the utilitarian SWF (the expected sum of individual well-being) and by EPP, a module for the prioritarian SWF (the expected sum of transformed well-being, here with a square-root transformation function). It is not satisfied by EEDEP, a different module for the prioritarian SWF (“EDE” is the equally-distributed-equivalent, here again with the square root transformation function; and EEDEP ranks policies according to the expected EDE).

Because the rank-weighted SWF violates Separability, no module for this SWF can satisfy Policy Separability. The table illustrates that one such module (the expected sum of rank-weighted well-being, with integer weights) violates Policy Separability.

28
Let’s now say that an individual is “sure to be unaffected” relative to the policy set $P$ if, for any pair of policies in the set, $P$ and $P^*$, that individual is sure to be unaffected between the two. In other words, an individual is “sure to be unaffected” relative to the policy set $P$ if his well-being is causally insulated from those policies. Whatever the state of nature may be, the decisionmaker’s selection of one rather than another policy in $P$ will have no causal impact on the individual’s welfare. Conversely, an individual is “possibly affected” relative to the policy set $P$ if he is not sure-to-be-unaffected.

If an SWF’s uncertainty module satisfies Policy Separability, then the population of ethical concern can be separated into two groups; those who are possibly affected by the policy choices at hand; and those who are sure to be unaffected. The decisionmaker can then “bracket” the well-being of the latter group. By this I mean that the decisionmaker need not incorporate their well-being into her calculations. To be clear, the sure-to-be-unaffected group remains part of the population of ethical concern. Their well-being is given the same weight as that of the possibly affected. But because the decisionmaker’s module satisfies Policy Separability, the ranking of policies is independent of the state-by-state well-being levels of those in the sure-to-be-unaffected group. Their well-being can be bracketed. More concretely, the decisionmaker does not need to assign probabilities to different possible well-being levels for each member of the sure-to-be-unaffected group, since her ranking of policies is the same for all such probability assignments.

By contrast, if the SWF’s uncertainty module fails Policy Separability, then the decisionmaker does need to assign probabilities to different possible well-being levels for each member of the sure-to-be-unaffected group. She also may need to consider how the different possible well-being levels of sure-to-be-unaffected individuals correlate with the well-being levels of the rest of the population (the possibly affected).

Being able to bracket the well-being of individuals who are sure-to-be-unaffected can be a great pragmatic advantage with respect to policy assessment. For example, imagine that the population of concern consists of the current population of a country, divided into twenty-six regions, A through Z. The policymaker is making a “local” policy choice, which (she is

---

39 I use the term “bracket” rather than “ignore” because the latter might suggest that the well-being of sure-to-be-unaffected individuals is given no weight. That is not the case: they remain fully part of the population of ethical concern.

40 Some uncertainty modules that fail Policy Separability require the decisionmaker not only to assign probabilities to different possible well-being levels of sure-to-be-unaffected individuals, but also to consider how those levels correlate with those of the possibly affected. This is true, for example, of the EEDEP module for prioritarianism (described below in Section 2.7 and illustrated in Table 2.6), or the module for a rank-weighted SWF that ranks policies according to the expected sum of rank-weighted well-being (illustrated in Table 2.6). Other uncertainty modules that fail Policy Separability are not sensitive to the correlation between the well-being levels of sure-to-be-unaffected and possibly affected individuals. This is true, for example, of a different module for a rank-weighted SWF: rank policies according to the sum of rank-weighted expected well-being.
confident) will only causally impact those living within Region A. Although unsure about the well-being levels of those living in Regions B through Z, she is reasonably certain that these levels (whatever they are) will not be altered by her policy choice. Those levels are fixed, i.e., causally independent of her policy choice: whatever the state of nature, the well-being levels of those living in Regions B through Z remain the same regardless of which policy the decisionmaker selects.

If the SWF’s uncertainty module fails Policy Separability, the policymaker needs to assign probabilities to different possible well-being levels for the various types of individuals living in Regions B through Z (the sure-to-be-unaffected), and perhaps also to consider how these levels would correlate with the well-being levels of those living in Region A (the possibly affected) as a result of each policy under consideration. By contrast, if the module satisfies Policy Separability, the policymaker can focus her modelling and information gathering on the individuals in Region A.

Policy Separability also facilitates the delegation of governmental choices to decisionmakers whose choices affect only a subset of the population. The regional example serves to illustrate this point. Imagine that a certain type of decision, type D*, directly impacts individuals living within a given region but has little effect on nonresidents; nonresidents are generally sure-to-be-unaffected with respect to decisions of type D*. (This might be the case, for example, regarding the level of public goods or governmental services that are provided only to residents, and that are funded by taxes borne by residents.) If so, and if a national policymaker endorses an SWF that satisfies Separability together with an uncertainty module that satisfies Policy Separability, the national policymaker can delegate decisions of type D* to a regional decisionmaker within each of the regions A through Z. The national policymaker can instruct each regional decisionmaker to make decisions of that type by using the chosen SWF and uncertainty module, and to “bracket” the well-being of individuals living outside the region. If the national policymaker’s preferred uncertainty module fails Policy Separability, delegation of this sort is not possible; rather, decisions of type D* within each region as well as other decisions must be made with reference to the well-being of the entire national population, affected or not.

A different type of example illustrating the advantages of Policy Separability involves prior generations. If the decisionmaker’s uncertainty module satisfies Policy Separability, she can bracket the well-being of individuals within the population of concern who are no longer alive at the time of her choice (“prior generations”). Dead individuals are quintessentially sure-to-be-unaffected; current choices don’t causally impact the well-being of the dead.

By contrast, how to handle prior generations creates a real dilemma for a decisionmaker with an uncertainty module that fails Policy Separability.41 (a) One option is to drop prior generations from the population of ethical concern. However, doing so can lead to problems of

---

41 For a discussion of this topic, see Fleurbaey (2010, pp. 663-69); Ferranna and Fleurbaey, chapter 8, this volume.
time-inconsistency. As individuals die, the population of concern changes. (b) The
decisionmaker can avoid time-inconsistency by including prior generations within the population
of concern. But this makes decisionmaking less tractable, since the well-being levels of these
dead individuals must now be considered.

The intention of the preceding paragraphs is to showcase the enhanced tractability of an
uncertainty module that satisfies Policy Separability, and thereby to argue in favor of the
Separability axiom. Again, satisfying the Separability axiom is a precondition (a necessary
condition) for an SWF to have a uncertainty module that conforms to Policy Separability. Rank-
weighted SWFs violate Separability, and thus are necessarily saddled with uncertainty modules
attended by the difficulties that result from violating Policy Separability.

Leximin and sufficientist SWFs do satisfy Separability, but are not score-based, and
hence are still not as tractable as generalized-utilitarian SWFs. It is quite economical to rank a
set of outcomes by using a formula to assign each a score, and then ordering the outcomes as per
the scores; leximin and sufficientist SWFs can’t do this. Further, a score-based SWF has a
natural, and highly tractable, uncertainty module: namely, taking the expected value of the score
(or of some function of the score). Uncertainty modules for the leximin and sufficientist SWFs
are more cumbersome. Finally, calculus techniques can be applied to the utilitarian scores and
(if the \( g(\cdot) \) function is chosen appropriately) to the prioritarian scores, so as to identify optimal
outcomes and policies. (Adler, 2019, p. 262 & n.6, 108-109, 279-80, 287-88). The fact that
leximin and sufficientist SWFs are not score-based is closely connected to their failure to satisfy
Continuity—since any SWF that satisfies this axiom is score-based. (Adler 2019, pp. 102-03.)

To be sure, tractability is only one normative dimension by which to evaluate SWFs. By
virtue of satisfying both Separability and Continuity, generalized-utilitarian SWFs are especially
well situated on this dimension. It’s certainly possible that other (more substantive)
considerations might weigh in favor of SWFs outside the generalized-utilitarian class.

But are there, indeed, strong (substantive) considerations that argue in favor of SWFs
outside the generalized-utilitarian class? Generalized utilitarians can plausibly deny that there
are. To begin, note that a concern for equity—for the fair distribution of well-being—can be
fully accomplished within the generalized-utilitarian class, by means of prioritarianism. Every
prioritarian SWF fully satisfies the Pigou-Dalton axiom.

Further, the prioritarian family of SWFs offers flexibility with respect to the degree of
concern for equity. Assume that one person (“Higher”) is better off than a second (“Lower”) in
vector \( v \). If \( w \) is reached from \( v \) via a pure gap-diminishing transfer of well-being from Higher
to Lower, then every prioritarian SWF prefers \( w \); this is what Pigou-Dalton requires. But what if
\( w \) is reached from \( v \) via a leaky (impure) gap-diminishing transfer of well-being between the
two: Higher’s well-being goes down by \( \Delta w \) and Lower’s well-being increases by a fraction
thereof, \( \rho \Delta w \), with \( 0 < \rho < 1 \). Pigou-Dalton doesn’t require a preference one way or the other, and indeed some prioritarian SWFs will prefer \( w \), while others will prefer \( v \).

More precisely, for any two well-being levels that Higher and Lower might have, and any degree of leakage in a gap-diminishing transfer between them, there are some prioritarian SWFs (with a transformation function sufficiently close to linear) that will disprefer the transfer, and other prioritarian SWFs (with a more concave transformation function) that will prefer the transfer.\(^{42}\) As the prioritarian transformation function becomes less concave, it exhibits a lesser degree of concern for equity, approaching utilitarianism at the limit. Conversely, as the prioritarian transformation function becomes more concave, it exhibits a greater degree of concern for equity, approaching leximin at the limit. (Bosmans, 2007).

In short, equity considerations don’t seem to furnish a strong reason for departing from generalized-utilitarian SWFs. Such SWFs (via prioritarianism) permit both a concern for equity, and flexibility regarding the degree of such concern. Nor do there appear to be other major considerations cutting against the generalized-utilitarian form. As we’ll see in detail in Section 2.7, prioritarianism under uncertainty has difficulty complying with the ex ante Pareto axiom. But this is also true of the leximin SWF, rank-weighted SWFs, and sufficientist SWFs. (Adler 2019, pp. 140-44).

Consider, now, the selection of an SWF within the generalized-utilitarian class. This chapter and book focus on the utilitarian SWF and prioritarian SWFs, ignoring generalized-utilitarian SWFs that are neither utilitarian nor prioritarian. Such SWFs have not been used much in economics, nor can they draw support from the philosophical literature. While the utilitarian SWF and prioritarian SWFs are the formal implementation of well-developed philosophical views (utilitarianism and prioritarianism, respectively), there appears to be virtually no philosophical argumentation in favor of a generalized-utilitarian approach that is neither utilitarian nor prioritarian.

Indeed, this absence of attention seems well justified. Generalized-utilitarian SWFs in this group, unlike prioritarian SWFs, can at most partly satisfy Pigou-Dalton (the equity axiom).\(^{43}\) Like prioritarian SWFs, but unlike the utilitarian SWF, these SWFs under uncertainty come into conflict with the ex ante Pareto axiom. (Adler 2019, pp. 140-44) So application under uncertainty is not grounds for choosing an SWF of this type, as opposed to a prioritarian SWF or

\(^{42}\) In particular, the degree of concavity of the Atkinson SWF, discussed in Section 2.9.1, is given by the parameter \( \gamma > 0 \). As \( \gamma \) increases, the Atkinson SWF becomes increasingly concave. For any gap-diminishing leaky transfer between a person at a given well-being level and a person at a lower well-being level, an Atkinson SWF with \( \gamma \) sufficiently small will disprefer the transfer, and an Atkinson SWF with \( \gamma \) sufficiently large will prefer it.

\(^{43}\) Recall that if an SWF is generalized-utilitarian (satisfying Pareto, Anonymity, Separability, and Continuity) and also satisfies Pigou-Dalton, then it is prioritarian. See note 37. A non-prioritarian, generalized-utilitarian SWF, \( \sum_{i=1}^{N} f(w_i) \), which is such that \( f() \) is strictly concave for some range of inputs would partly satisfy Pigou-Dalton—for transfers between individuals with well-being levels in that range.
the utilitarian SWF. Finally, they are no more tractable than prioritarian SWFs or the utilitarian SWF. Thus it is hard to see what would warrant their adoption.

The normative question that remains to be engaged is the debate between utilitarianism and prioritarianism. That will be undertaken in Section 2.8.

2.5 Utility, Prioritarianism and Interpersonal Comparisons

Some economists are skeptical about interpersonal well-being comparisons. However, the SWF methodology (including, in particular, the utilitarian SWF and prioritarian SWFs) requires some degree of interpersonal comparability. On SWFs and interpersonal comparability, see Adler (2019, ch. 2 and appendix); Bossert and Weymark (2004).

Assume that only intrapersonal well-being comparisons are meaningful. Let’s refer to this position as No Interpersonal Comparisons. One, especially stark, version of No Interpersonal Comparisons says: only intrapersonal well-being level comparisons are meaningful. A different, less severe version, allows for both level and difference comparisons—but only intrapersonally, not interpersonally.

In order to explain the implications of No Interpersonal Comparisons for the SWF framework, I’ll need to introduce the concept of an “individual-specific cardinal rescaling” of a well-being measure. Take a well-being measure \( w(\cdot) \). Now define a new well-being measure \( w^*(\cdot) \) as follows. For a given individual \( i \), choose individual-specific constants \( k_i > 0 \) and \( l_i \). Then our new well-being measure is the old measure rescaled by these constants. That is, for every outcome \( x \), \( w^*_i(x) = k_i \times w_i(x) + l_i \).

If No Interpersonal Comparisons is true, then the two well-being measures are informationally equivalent. They contain the very same meaningful well-being information—meaningful according to the view at hand. Assume, first, that only intrapersonal level comparisons are meaningful. Note that \( w^*(\cdot) \) contains the very same information as \( w(\cdot) \) regarding these comparisons: whenever \( w_i(x) \geq w_i(y) \), then \( w^*_i(x) \geq w^*_i(y) \), and vice versa. Assume, next, that only intrapersonal level and difference comparisons are meaningful. Again, \( w^*(\cdot) \) contains the very same information as \( w(\cdot) \) regarding these comparisons. (1) Whenever \( w_i(x) \geq w_i(y) \), then \( w^*_i(x) \geq w^*_i(y) \), and vice versa. (2) Whenever \( w_i(x) - w_i(y) \geq w_i(z) - w_i(zz) \), \( w^*_i(x) - w^*_i(y) \geq w^*_i(z) - w^*_i(zz) \) and vice versa.

Observe, next, that our chosen SWF (whatever it might be) should be invariant to swapping a well-being measure for an informationally equivalent alternative. Assume that \( w(\cdot) \) and \( w^*(\cdot) \) are informationally equivalent. Imagine, now, that our chosen SWF ranks outcomes one way using \( w(\cdot) \) and a different way using \( w^*(\cdot) \). This would be quite problematic. If \( w(\cdot) \) and \( w^*(\cdot) \) are indeed informationally equivalent, they differ only numerically—only in terms of which numbers are being used to represent the very same well-being facts. Our chosen SWF, whatever it might, shouldn’t be sensitive to this merely numerical difference. Welfarists want
outcomes to be ranked in light of their well-being patterns—and, if \( w(\cdot) \) and \( w*(\cdot) \) are indeed informationally equivalent, the patterns of well-being in all the outcomes as signaled by \( w(\cdot) \) is the same as those patterns as signaled by \( w*(\cdot) \).

Putting the last two paragraphs together, we arrive at the following conclusion. If No Interpersonal Comparisons is true, our chosen SWF (whatever it might be) should be invariant to swapping any given well-being measure \( w(\cdot) \) for an individual-specific cardinal rescaling of \( w(\cdot) \). But now a deep problem arise. A basic theorem of the SWF literature shows that no non-dictatorial Paretian SWF is invariant to individual-specific cardinal rescalings of the well-being measure. (Adler 2019, pp. 45-46; Bossert and Weymark, 2004, p. 1146; Weymark, 2016, p. 152). Yet any plausible SWF, by welfare-consequentialist lights, will be both non-dictatorial (this is an even weaker impartiality requirement than Anonymity) and Paretian. In short, no plausible SWF is invariant to individual-specific cardinal rescalings of the well-being measure. Every such SWF presupposes that well-being numbers contain more information than merely information about intrapersonal level and difference comparisons.

Table 2.7 illustrates specifically that neither the utilitarian SWF nor a prioritarian SWF, nor the other major types of SWFs discussed in this chapter (leximin, rank-weighted, and sufficientist) will be invariant to individual-specific cardinal rescalings of the well-being measure.45

---

44The Cobb-Douglas SWF (the product of well-being numbers, assumed to be positive) is invariant to individual-specific ratio rescalings of well-being numbers. This SWF ranks well-being vectors the same way as the sum of the logarithm of well-being; in short, this is a type of prioritarian SWF. No other SWF that satisfies Strong Pareto and Anonymity is invariant to individual-specific ratio rescalings. (Bossert and Weymark 2004, p. 1148). However, an individual-specific ratio rescaling does preserve certain interpersonal information, namely whether individuals are above or below the zero point. Thus it is not correct to say that the Cobb-Douglas SWF is consistent with No Interpersonal Comparisons.

45The example in Table 2.7, and the theorem mentioned in the previous paragraph, assume a profile-independent SWF. An SWF is profile-independent if its functional form does not depend upon \( w(\cdot) \). The very same rule is used to rank well-being vectors, and thereby to rank outcomes via the Master Recipe, regardless of which well-being measure \( w(\cdot) \) is used to map outcomes onto vectors. See Section 2.9.1. The utilitarian SWF is necessarily profile-independent; there is just one SWF here, namely the sum of well-being numbers. Profile-dependence becomes a possibility within the family of prioritarian SWFs; the transformation function \( g(\cdot) \) might vary as \( w(\cdot) \) does. This in turn raises the question whether No Interpersonal Comparisons might be consistent with a profile-dependent prioritarian SWF.

The question will not be pursued. Profile-independence is theoretically preferred; see Section 2.9.1. Thus, the fact that both the utilitarian SWF and any profile-independent prioritarian SWF are inconsistent with No Interpersonal Comparisons makes a strong case for interpersonal comparability.
Table 2.7: An individual-specific cardinal rescaling of the well-being measure

<table>
<thead>
<tr>
<th>Original well-being numbers</th>
<th>Individual-specific cardinal rescaling</th>
<th>Scaling factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Outcome</td>
<td>x</td>
</tr>
<tr>
<td>Abel</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Bob</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>Chloe</td>
<td></td>
<td>49</td>
</tr>
<tr>
<td>Diana</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Util. score</td>
<td></td>
<td>84</td>
</tr>
<tr>
<td>Prior. score(√)</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Rk-wtd score (integer)</td>
<td></td>
<td>130</td>
</tr>
<tr>
<td>Leximin</td>
<td>z &gt; y &gt; x</td>
<td></td>
</tr>
<tr>
<td>Sufficientist (✓) and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(w_{threshold} = 100)</td>
<td>z &gt; y &gt; x</td>
<td></td>
</tr>
</tbody>
</table>

*Explanation:* The left of the table displays well-being vectors assigned to outcomes \(x\), \(y\), and \(z\) by a well-being measure \(w(\cdot)\), while the right displays vectors per an individual-specific cardinal rescaling of \(w(\cdot)\). The table illustrates that neither the utilitarian SWF, nor a prioritarian SWF, nor a rank-weighted SWF, nor the leximin SWF, nor a sufficientist SWF is invariant to an individual-specific cardinal rescaling of well-being numbers. The square root transformation function is used for the prioritarian and sufficientist SWFs; and the rank-weighted SWF uses integer weights.

I’ve mentioned that a particular version of No Interpersonal Comparisons is especially stark—allowing only intrapersonal level comparisons. Some economists are suspicious of difference comparisons, and that is reflected in this severe view. It says, no difference comparisons of any type are meaningful, neither intra- nor interpersonal; intrapersonal level comparisons are meaningful; interpersonal level comparisons are not.

There is a view regarding well-being comparability that relaxes skepticism regarding interpersonal comparisons, but retains skepticism about difference comparisons. According to such a view, both intrapersonal and interpersonal level comparisons are meaningful, but neither intrapersonal nor interpersonal difference comparisons are.

Call such a view No Difference Comparisons. As illustrated by Table 2.8, the proponent of the leximin SWF can accept No Difference Comparisons without difficulty. But both the utilitarian SWF and prioritarian SWFs are inconsistent with No Difference Comparisons.
Table 2.8: A common ordinal rescaling of the well-being measure

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Original well-being numbers</th>
<th>Common ordinal rescaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abel</td>
<td>1, 4, 9</td>
<td>Abel 100, 101, 102</td>
</tr>
<tr>
<td>Bob</td>
<td>36, 25, 16</td>
<td>Bob 110, 104, 103</td>
</tr>
<tr>
<td>Chloe</td>
<td>36, 9, 16</td>
<td>Chloe 110, 102, 103</td>
</tr>
<tr>
<td>Diana</td>
<td>1, 25, 16</td>
<td>Diana 100, 104, 103</td>
</tr>
<tr>
<td>Util. score</td>
<td>74, 63, 57</td>
<td>420, 411, 411</td>
</tr>
<tr>
<td>Prior. score(√)</td>
<td>14, 15, 15</td>
<td>40.976, 40.545, 40.546</td>
</tr>
<tr>
<td>Leximin</td>
<td>z &gt; y &gt; x</td>
<td>z &gt; y &gt; x</td>
</tr>
</tbody>
</table>

Explanation: The left of the table displays well-being vectors assigned to outcomes x, y, and z by a well-being measure \( w(\cdot) \), while the right displays vectors assigned by a common ordinal rescaling of \( w(\cdot) \). The well-being numbers 1, 4, 9, 16, 25, and 36 are rescaled as 100, 101, 102, 103, 104 and 110 respectively. Common ordinal rescalings (see note __ for a definition) preserve intra- and interpersonal well-being level information but not well-being difference information. As illustrated here, the leximin SWF is invariant to a common ordinal rescaling of the well-being measure, but the utilitarian SWF and prioritarian SWFs are not.

In short, anyone who endorses the utilitarian SWF or prioritarian SWFs must reject No Interpersonal Comparisons, and she must reject No Difference Comparisons. Our theory of well-being (whether hedonic, preference-based, or objective-good) should be such as to allow for both intra- and interpersonal comparisons, of both levels and difference. It should count all such comparisons as meaningful. And all four types of comparisons should be reflected in the well-being measure \( w(\cdot) \).

Indeed, making all four types of comparisons is common sense. No Interpersonal Comparisons and No Difference Comparisons each fly in the face of common sense.

---

46 Three possible views regarding intra- and interpersonal comparisons of well-being levels and differences have been considered and rejected here: that only intrapersonal level comparisons are meaningful; that only intrapersonal level and difference comparisons are meaningful; and that only intra- and interpersonal level comparisons are meaningful. Logically, rejecting these three views doesn’t entail accepting that all four types of comparisons are meaningful. There are additional possibilities: for example, that only interpersonal comparisons are meaningful; that only intra- and interpersonal difference comparisons are meaningful; etc. However, these further possibilities can be rejected out of hand; none are even superficially plausible.
Imagine that John, in one possible world, is impoverished, has no friends, no job, is in poor health, and experiences chronic pain. In a second possible world, John has a good income, many friends, an engaging job, excellent health, and rarely feels pain. Then (ceteris paribus) John in the second world is better off than John in the first. This is an intrapersonal level comparison; no one disputes that such a comparison is meaningful.

But imagine now that John, in actuality, has a good income, many friends, an engaging job, excellent health, and rarely feels pain; while a different person, Jane, is impoverished, friendless, unemployed, in poor health, and experiences chronic pain. Surely John is better off than Jane (an interpersonal level comparison), just as John in the second world is better off than John in the first.

Difference comparisons are also common sense. Two days ago, Audra woke up with a slightly itchy elbow, but relieved the itch with a cortisone cream. Yesterday, she woke up with a terrible migraine headache, but it stopped after she took a pill. Intuitively, the headache pill made a bigger difference for Audra’s well-being than the cortisone cream (an intrapersonal difference comparison). To insist that the two changes in Audra’s well-being are incomparable (as does No Difference Comparisons) is counterintuitive.

Nor is it plausible to allow for intrapersonal difference comparisons, but reject interpersonal difference comparisons. Tweak the Audra case so that Audra, two days ago, wakes up with an itchy elbow while a different person, Brian, wakes up yesterday with the migraine. Intuitively, the headache pill made a bigger difference for Brian’s well-being than the cream did for Audra’s (an interpersonal difference comparison).

To reiterate: Our theory of well-being (whether hedonic, preference-based, or objective-good) should be such as to allow for both intra- and interpersonal comparisons, of both levels and differences. And all four types of comparisons should be reflected in our well-being measure \( w(\cdot) \).

Through this chapter, I will assume that \( w(\cdot) \) does indeed represent both intra- and interpersonal well-being comparisons, of both levels and difference. How to construct such a \( w(\cdot) \) is the topic of Chapter 3.

### 2.6 Is Prioritarianism Genuinely Different from Utilitarianism?

The prioritarian rule for ranking well-being vectors is different from the utilitarian rule. Prioritarianism assigns each vector \( w \) a score equaling \( g(w_1) + g(w_2) + \ldots + g(w_N) \) and ranks the vectors in the order of these scores; while utilitarianism assigns each vector \( w \) a score equaling \( w_1 + w_2 + \ldots + w_N \) and ranks the vectors in the order of these scores.

These rules are also axiomatically distinct. As explained above, the prioritarian vector-ranking rule satisfies the Pigou-Dalton axiom, while the utilitarian rule does not.
Finally, the prioritarian recipe for ranking outcomes is different from the utilitarian recipe. The prioritarian recipe is: outcome \( x \) at least as good as \( y \) iff \( g(w_1(x)) + \ldots + g(w_N(x)) \geq g(w_1(y)) + \ldots + g(w_N(y)) \). The utilitarian recipe is: outcome \( x \) at least as good as \( y \) iff \( w_1(x) + \ldots + w_N(x) \geq w_1(y) + \ldots + w_N(y) \).

Still, the skeptic might argue that utilitarianism and prioritarianism are not genuinely distinct. Why not? The prioritarian ranking of outcomes that results from combining well-being measure \( w(\cdot) \) and transformation function \( g(\cdot) \) is identical to the utilitarian outcome ranking using a different well-being measure, \( w^*(\cdot) \). Simply define \( w^*(\cdot) \) as follows: \( w^*_i(x) = g(w_i(x)) \) for every outcome \( x \). In other words, our supposedly “prioritarian” outcome ranking can be expressed as a utilitarian outcome ranking by taking the transformed well-being numbers that result from our transformation function and initial well-being measure \( w(\cdot) \), and relabeling them as straight well-being numbers according to a new well-being measure \( w^*(\cdot) \).

The following example illustrates the skeptic’s argument. Imagine that individuals are modelled as having two attributes, income and health quality, the latter quantified on a 0-1 scale (1 indicating perfect health, 0 a health state no better than death). Our well-being measure \( w(\cdot) \) is defined as the product of health quality and the logarithm of income. Suppose now that we pick the square root function as the prioritarian transformation function. The “prioritarian” score assigned to each outcome is therefore calculated as follows: for each individual, multiply her health quality by the logarithm of her income to determine her well-being; take the square root of this well-being number to arrive at her transformed well-being; sum up these transformed-well-being numbers. But this “prioritarian” score is nothing other than a utilitarian score calculated with a new well-being measure \( w^*(\cdot) \) defined as follows: well-being is the square root of the product of health quality and the logarithm of income.

It is no doubt true that prioritarianism with a given well-being measure and transformation function is identical—in its outcome ranking—to utilitarianism with a different well-being measure. But the skeptic is wrong to conclude that prioritarianism is “the same” as utilitarianism. (1) Prioritarianism and utilitarianism are, obviously, distinct relative to a given well-being measure. If \( w(\cdot) \) is our well-being measure, then the outcome ranking achieved using the prioritarian recipe is different from the ranking achieved using the utilitarian recipe. (2) It is true that, by combining \( w(\cdot) \) and \( g(\cdot) \), we can arrive at a new well-being measure \( w^*(\cdot) \), such that the utilitarian outcome ranking with well-being measured by \( w^*(\cdot) \) is the same as the prioritarian ranking with well-being measured by \( w(\cdot) \) and transformed by \( g(\cdot) \). But \( w(\cdot) \) and \( w^*(\cdot) \) are different well-being measures. In particular, intra- and interpersonal difference comparisons, as per \( w^*(\cdot) \), are different from intra- and interpersonal difference comparisons, as per \( w(\cdot) \). Our theory of well-being, we have supposed, allows for difference comparisons as well as level comparisons. If \( w(\cdot) \) is an eligible measure—eligible according to our theory, in the sense of

---

accurately representing the theory’s level and difference comparisons—then \( w^*(\cdot) \) is not an eligible measure.

The discussion here illustrates a more general point about the SWF framework—namely, that this decision-procedure is comprised of distinct components, meant to capture separable and distinctive aspects of ethical reasoning. One is the well-being component. We deliberate about what makes an individual life go well or badly, and thereby arrive at a theory of well-being, mathematically represented via a well-being measure \( w(\cdot) \). A second, separate component involves the specification of an SWF. Here, we deliberate about how the ethical goodness of outcomes depends upon the pattern of individual well-being. The question at this stage of ethical deliberation is not the nature of well-being, but its ethical role. Why would well-being vector \( w \) be better, worse, or equally good as well-being vector \( v \)? Should we endorse axioms such as Pareto, Anonymity, Pigou-Dalton, Separability, and Continuity? What is the specific form of the vector-ranking rule, \( \succeq \)? If we conflate the two components, it may look like prioritarianism and utilitarianism are the same. But if we are sensitive to the difference between debates about well-being and debates about the ethical role of well-being, the two SWFs are—clearly—different. Having settled upon a well-being measure \( w(\cdot) \), it is a meaningful—and important!—question whether the ethical comparison of outcomes depends upon the simple sum of well-being, or instead the sum of transformed well-being.

2.7 Prioritarianism and Utilitarianism under Uncertainty

Recall that we are using the “state of nature” setup (Section 2.2.4) to conceptualize choice under uncertainty. Any particular policy that government might select can be represented as a state-by-state array of outcomes. With a well-being measure in hand, each policy corresponds to a state-by-state array of well-being vectors. Any given SWF \( \succeq \) has various possible uncertainty “modules.” An uncertainty module is some rule for ranking policies, understood as state-by-state arrays of well-being vectors. Each such module satisfies the basic constraint that it ranks degenerate policies consistently with the SWF.

Let’s begin with prioritarianism. There is a substantial body of scholarship on prioritarianism under uncertainty.49 The most widely discussed prioritarian uncertainty modules are the following three: (1) “ex post prioritarianism” (EPP), namely ranking policies according to the expected sum of transformed well-being; (2) “ex ante prioritarianism” (EAP), namely

---

48 \( w^*(\cdot) \) is a common ordinal rescaling of \( w(\cdot) \) if there exists a strictly increasing function \( f(\cdot) \) such that \( w^*_i(x) = f(w_i(x)) \) for all \( i, x \). \( w^*(\cdot) \) is a common cardinal rescaling of \( w(\cdot) \) if there exists \( k > 0, l \), such that \( w^*_i(x) = k \times w_i(x) + l \) for all \( i, x \). Note that if \( w^*(\cdot) \) is a common cardinal rescaling of \( w(\cdot) \), then it is a common ordinal rescaling; but the converse is not true. Note also that if \( w^*(\cdot) \) is a common ordinal rescaling of \( w(\cdot) \) but not a common cardinal rescaling, \( w^*(\cdot) \) will preserve the intra- and interpersonal level comparisons of \( w(\cdot) \) but need not preserve the intra- and interpersonal difference comparisons of \( w(\cdot) \).

\( w^*(\cdot) \) defined as \( w^*(x) = g(w_i(x)) \), \( g(\cdot) \) strictly increasing and strictly concave, is a common ordinal but not common cardinal rescaling of \( w(\cdot) \)—hence need not preserve the difference information in \( w(\cdot) \).

49 That scholarship is reviewed in Adler (2012, ch. 7; 2019, ch. 3-4); Adler & Holtug (2019). See also Fleurbaey (2010); Mongin and Pivato (2016).
ranking policies according to the sum of transformed expected well-being; and (3) expected equally-distributed-equivalent prioritarianism (EEDEP), namely ranking policies according to the expected value of the equally-distributed-equivalent of the well-being pattern in each state. See Appendix for formulas.

One might wonder why the literature hasn’t settled on a single, canonical version of prioritarianism under uncertainty. Why the debate? The answer is axiomatic. Just as we posited axioms for the ranking of well-being vectors (Pareto, Anonymity, Pigou-Dalton, etc.), so we might posit axioms for the ranking of policies—for short, “uncertainty axioms.” There are a variety of plausible uncertainty axioms, but I’ll focus here on three: ex ante Pareto, Dominance, and Policy Separability (the last already mentioned). Each of the three seems quite plausible, and yet it is impossible for a prioritarian uncertainty module to satisfy all three. The debate between EPP, EAP, and EEDEP is fueled by disagreement about which of the three uncertainty axioms to drop given that all three can’t be satisfied.

The uncertainty axioms are as follows. (See Appendix for precise statement.)

**Ex Ante Pareto.** (1) **Ex Ante Pareto Indifference.** If each person’s expected well-being with policy $P^*$ is the same as her expected well-being with policy $P$, then the two policies are equally good. (2) **Ex Ante Strong Pareto.** If each person’s expected well-being with policy $P^*$ is greater than or equal to her expected well-being with policy $P$, and at least one person’s expected well-being with $P^*$ is strictly greater than her expected well-being with $P$, then $P^*$ is better than $P$.

**Dominance.** Assume that in each state of nature, the well-being vector produced by policy $P^*$ is better than the well-being vector produced by policy $P$—better according to some given SWF. Then the uncertainty module for that SWF should be such as to rank $P^*$ better than $P$.

**Policy Separability.** Assume that some individuals are “sure to be unaffected” as between two policies $P$ and $P^*$ (meaning that each such individual with $P$ has the same well-being level in any given state that she does with $P^*$). If so, the ranking of the two policies should be invariant to the state-by-state well-being levels of these “sure to be unaffected” individuals.

Table 2.9 illustrates the fundamental conflict between ex ante Pareto and Dominance, for a prioritarian SWF. As the table shows, it is impossible for an uncertainty module for a prioritarian SWF to satisfy both ex ante Pareto and Dominance.
Table 2.9: The conflict between ex ante Pareto and Dominance

<table>
<thead>
<tr>
<th>State $s$</th>
<th>Expected $\pi(s) = .5$</th>
<th>State $s^*$</th>
<th>Expected $\pi(s^*) = .5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jasper</td>
<td>90 10 50</td>
<td>50 $-\varepsilon$</td>
<td>50 $-\varepsilon$</td>
</tr>
<tr>
<td>Karl</td>
<td>10 90 50</td>
<td>50 $-\varepsilon$</td>
<td>50 $-\varepsilon$</td>
</tr>
</tbody>
</table>

Explanation: Dominance requires the uncertainty module for a prioritarian SWF to rank $P^*$ over $P$ for some $\varepsilon > 0$. However, ex ante Strong Pareto requires that $P$ be preferred to $P^*$. Setting $\varepsilon = 0$ illustrates the conflict between Dominance and ex ante Pareto Indifference.

The EAP uncertainty module satisfies ex ante Pareto, but at the cost of violating Dominance. Conversely, the EPP and EEDEP uncertainty modules satisfy Dominance, but at the cost of violating ex ante Pareto.

As illustrated by Table 2.9, the prioritarian is forced to make an axiomatic choice. Ex ante Pareto and Dominance are each intuitively plausible, and yet can’t be jointly satisfied (given a prioritarian $\succ$). Thus the prioritarian needs to consider relative plausibility. Which of these competing uncertainty axioms—ex ante Pareto or Dominance—is more compelling?

This question of relative plausibility is, of course, debatable, but I believe a strong case can be made in favor of Dominance rather than ex ante Pareto. Dominance seems to flow directly from consequentialism. Any welfare-consequentialist agrees that the outcome ranking drives policy choice under conditions of certainty. In particular, and most fundamentally: If (1) we know for sure that policy $P$ will lead to outcome $x$, with well-being pattern $w$, and that policy $P^*$ will lead to outcome $x^*$, with well-being pattern $w^*$; and (2) our ranking of well-being patterns is such that $w^*$ is better than $w$, so that $x^*$ is a better outcome than $x$, then (3) we should conclude that $P^*$ is a better policy than $P$. (To accept this axiom seems definitonal of welfare-consequentialism.) But it is very hard to justify accepting this axiom, yet rejecting Dominance. If $P^*$ and $P$ are as described by the Dominance axiom, we may not know for sure what outcomes and attendant well-being patterns (vectors) will result from the two policies; but what we do know is that, in each and every state, the $P^*$ pattern (vector) is better than the $P$ pattern.

Dominance can also be supported by appeal to an ideal, fully-informed adviser. If $P^*$ and $P$ are as described by the Dominance axiom, then the decisionmaker can be sure that a fully informed adviser would recommend $P^*$. (Fleurbaey 2010, p. 655; Fleurbaey and Voorhoeve 2013).

Let’s now compare EPP and EEDEP. Again, both of these uncertainty modules for the prioritarian SWF satisfy Dominance. How do they differ axiomatically?
While both approaches violate the ex ante Pareto axiom, EEDEP satisfies these axioms in a restricted sense. In a nutshell, EEDEP satisfies the ex ante Pareto axiom when individuals are identically situated. (More precisely: if the well-being vector produced by policy $P$ in each state is such that all individuals have the same well-being level, and the well-being vector produced by policy $P^*$ in each state is such that all individuals have the same well-being level, then: the $P/P^*$ ranking by EEDEP satisfies the ex ante Pareto axiom.) Conversely, EPP violates the ex ante Pareto axiom even when individuals are identically situated. However, EPP satisfies Policy Separability while EEDEP violates Policy Separability. (See Table 2.6.)

In my own view, EPP is on balance better justified than EEDEP. Consider, first, the axiom of Policy Separability. The pragmatic advantages of Policy Separability are large, as discussed earlier (Section 2.4). EEDEP can be criticized for failing to realize the potential pragmatic benefits of Separability—the pragmatic benefits that occur if the separable SWF is conjoined with a module that satisfies Policy Separability (as in the case of EPP). Further, as explained in the notes, it can be argued that the combination of a separable SWF with an uncertainty module that violates Policy Separability violates a maxim of rational choice—namely, don’t rank choices in a manner that you know a fully informed adviser would disapprove.50

On the other hand, EEDEP satisfies ex ante Pareto when individuals are identically situated. This is an axiomatic benefit, to be sure, but—arguably—not sufficient to justify dropping Policy Separability. The supposition that individuals are identically situated is quite restrictive. “Identically situated” means everyone in the entire population of ethical concern—not merely those in a subpopulation. Consider again the example of Regions A through Z (Section 2.4), with policies that affect only those in the first region. Tweak the example so that everyone in Region A is identically situated. (Prior policies and practices, let’s imagine, have operated to produce an equal distribution of well-being within this region.). Now imagine a local policy choice that affects only those living in Region A. Assume that the expected well-being of those in region A is greater with policy $P^*$ than with policy $P$. If so, the ex ante Pareto axiom requires that $P^*$ be ranked above $P$. EPP can violate ex ante Pareto in this case. But EEDEP can also violate ex ante Pareto in this case.

50 Assume that Policy Separability applies and requires the $P/P^*$ ranking to be the same as the $P^*/P^{**}$ ranking. That is, there are some sure-to-be-unaffected individuals each of whom has the same well-being in any given state $s$ with $P$ as she does with $P^*$. These individuals are also sure-to-be unaffected as between $P^*$ and $P^{**}$. And the state-by-state pattern of well-being for every other individual is the same in the $P/P^*$ pair as in the $P^*/P^{**}$ pair.

For any given state $s$, let $w(P, s)$ denote the well-being vector that results in state $s$ with a given policy $P$. Because the SWF is separable, it must be the case that, for every state $s$, $w(P, s) \succeq w(P^*, s)$ iff $w(P^*, s) \succeq w(P^{**}, s)$. (This is illustrated in Table 2.6.) A fully informed adviser will know what the state of nature is, and will rank $P$ at least as good as $P^*$ iff the adviser ranks $P^*$ at least as good as $P^{**}$. The imperfectly informed decisionmaker doesn’t herself know what the state of nature is; but, by the structure of the problem, she knows that the adviser will never advise her to rank the $P/P^*$ pair differently than the $P^*/P^{**}$ pair.
More generally, EEDEP (like EPP) can violate ex ante Pareto whenever some individuals in the population of ethical concern are differently situated from other individuals in that entire population. Even if the population can be partitioned into two subgroups, individuals identically situated (within the subgroup) and individuals sure to be unaffected—as in the example immediately above—EEDEP (like EPP) may violate ex ante Pareto.\footnote{Adler (2019, pp. 130-138, 277-82) describes ex ante Pareto limited to cases in which individuals are either identically situated or sure-to-be-unaffected as “heartland ex ante Pareto,” and discusses the fact that any prioritarian uncertainty module which satisfies Dominance violates heartland ex ante Pareto.}

I’ve discussed prioritarianism under uncertainty at some length. The specification of an uncertainty module for the prioritarian SWF is not straightforward. EAP violates Dominance, but satisfies ex ante Pareto (and, as it happens, Policy Separability). EPP violates ex ante Pareto, but satisfies Dominance and Policy Separability. EEDEP satisfies Dominance; satisfies ex ante Pareto in the special case when individuals (in the entire population of ethical concern) are identically situated; but violates Policy Separability. I believe EPP to be on balance best justified, but this is certainly a matter for debate.

By contrast, the specification of an uncertainty module for the utilitarian SWF is straightforward. The module generally adopted in the literature (for short, UUU—utilitarianism under uncertainty) ranks policies according to the expected sum of individual well-being. This module satisfies Dominance and ex ante Pareto and Policy Separability. If the SWF is utilitarian, there exists an uncertainty module—UUU—that satisfies all three of these plausible uncertainty axioms.

2.8 Utilitarianism versus Prioritarianism

We now turn to the debate between utilitarianism and prioritarianism. To facilitate this debate, Table 2.10 displays how both the utilitarian SWF and prioritarian SWFs fare with respect to the SWF axioms discussed earlier.

The table adds a further axiom, Minimal Pigou-Dalton, which states the following.

**Minimal Pigou-Dalton.** There exists at least one pair of well-being vectors \( w \) and \( v \) such that \( w > v \) and such that the two vectors are as follows: \( i \) is better off than \( j \) in vector \( v \); in \( w \) \( i \)’s well-being has been decreased by \( \Delta w \), while \( j \)’s well-being has increased by the very same amount, and the gap between the two individuals has decreased; and every other person is equally well off in \( w \) as she is in \( v \).

In other words, while the Pigou-Dalton axiom articulates a universal constraint, Minimal Pigou-Dalton articulates the corresponding existential constraint. Pigou-Dalton says: whenever vector \( w \) is reached from vector \( v \) via a pure, gap-diminishing transfer of well-being from someone better off to someone worse off, affecting no one else, \( w \) is better than \( v \). Minimal Pigou-Dalton says: there exists at least one case in which vector \( w \) is reached from vector \( v \) via a pure, gap-
diminishing transfer of well-being from someone better off, to someone worse off, affecting no one else, and \( w \) is better than \( v \). Note that Pigou-Dalton implies Minimal Pigou-Dalton, but not vice versa.\(^{52}\)

In assessing the relative merits of the utilitarian SWF as compared to prioritarian SWFs, it’s also important to consider the properties of the possible uncertainty modules associated with each. Here, I’ll focus on the utilitarian SWF as implemented under uncertainty via the UUU module, as compared to the prioritarian SWF as implemented under uncertainty via the EPP module. (The reader who prefers EEDEP or EAP to EPP as a prioritarian procedure for handling uncertainty can undertake a similar exercise.) For the remainder of this Section, I’ll use the terms “utilitarianism” or “the utilitarian SWF” to mean, specifically, the utilitarian SWF together with the UUU module; and I’ll use the terms “prioritarianism” or “a prioritarian SWF” to mean, specifically, a prioritarian SWF together with the EPP module.

Table 2.10 also displays how the utilitarian SWF and prioritarian SWFs fare with respect to the uncertainty axioms discussed earlier.

**Table 2.10: The axiomatic properties of utilitarian and prioritarian SWFs**

<table>
<thead>
<tr>
<th></th>
<th>Utilitarian SWF</th>
<th>Prioritarian SWFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separability</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Continuity</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pigou-Dalton</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Minimal Pigou-Dalton</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>with UUU module</td>
<td>with EPP module</td>
</tr>
<tr>
<td>Policy Separability</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Dominance</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ex Ante Pareto</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

As the table illustrates, utilitarianism and prioritarianism are alike in many respects. The features that distinguish them are twofold: ex ante Pareto and equity (Pigou-Dalton and Minimal Pigou-Dalton).

That prioritarianism violates ex ante Pareto is, no doubt, an intuitively problematic feature of the approach. (Again, we are focusing here on prioritarianism applied under uncertainty with EPP.)\(^{53}\) It seems perverse for it to be the case that everyone has the same

---

\(^{52}\) That is, if an SWF satisfies Pigou-Dalton, it satisfies Minimal Pigou-Dalton. However, it is possible for an SWF to satisfy Minimal Pigou-Dalton but not Pigou-Dalton. (Indeed, this is true of sufficientist SWFs.)

\(^{53}\) Recall that Prioritarianism applied under uncertainty with EEDEP also violates ex ante Pareto, except when individuals are identically situated; prioritarianism applied with EAP satisfies ex ante Pareto, but at the cost of violating Dominance.
expected well-being with policy $P^*$ as with policy $P$, and yet the policies are not ranked equal. It seems yet more perverse that (a) some or even all individuals have greater expected well-being with policy $P^{++}$ than with policy $P^*$, and (b) no one has lower expected well-being with policy $P^{++}$, and yet $P^{++}$ is not ranked better than $P^*$.

But we should remind ourselves that the prioritarian SWF does conform to the Pareto axiom, which is a constraint on the outcome ranking. Relatedly, the prioritarian SWF does satisfy ex ante Pareto with respect to degenerate policies. If the decisionmaker knows for sure that policy $P$ yields vector $w$, and that policy $P^*$ yields vector $w^*$, then she will rank $P^*$ better than/equal to $P$ if ex ante Pareto requires her to do so.\footnote{Moreover, if the well-being vector produced by $P^*$ in each state is required by the Pareto axiom to be ranked better than/equally good as the well-being vector produced by $P$ in that state, then prioritarianism will rank $P^*$ better than/equally good as $P$.}

Conversely, the failure of the prioritarian SWF to satisfy ex ante Pareto arises only by virtue of the decisionmaker’s imperfect information. In the non-degenerate (imperfect information) case, the decisionmaker needs to compare $P$ to $P^*$ and yet does not know, for each person, what his well-being level would be were $P$ to be chosen and what it would be were $P^*$ to be chosen instead. Ex ante Pareto “extends” the Pareto idea to this imperfect-information context by calculating well-being expectations for each person. That extension seems intuitive, but also can have troubling results. The ex ante Pareto axiom can require that $P^*$ be ranked above $P$ even though, in fact—given the actual state of nature—some will be worse off with $P^*$. Indeed, the ex ante Pareto axiom can require that $P^*$ be ranked above $P$ even though, in fact, everyone will be worse off with $P^*$.

Let’s now turn to the major weakness of utilitarianism, namely its failure to satisfy equity axioms.

In other work, I have offered a systematic defense of the Pigou-Dalton axiom, via the notion of “claims.” (Adler 2012, ch. 5) Imagine that we justify the ranking of any two well-being vectors (patterns of well-being) by considering that comparison from the perspective of each individual. Specifically, as between $w$ and $v$, a given person has a claim in favor of $w$ over $v$ if she is better off in $w$ than $v$; a claim in favor of $v$ over $w$ if she is better off in $v$ than $w$; and no claim if she is equally well off in the two vectors. This claim-based justificatory framework tries to give specific content to the somewhat amorphous notion (introduced by John Rawls) that ethics should take seriously the “separateness of persons.” We do so, according to this framework, by proceeding person-by-person and, for each one, considering the direction and strength of her claim between the two possible well-being vectors.

This claims framework, I argue, reveals a deep unity between the Pareto and Pigou-Dalton axioms. It immediately supports Pareto. (If everyone is equally well off in two vectors, then none have claims between them, and so it follows that the two are equally good. If some
are better off in \( w \) than \( v \), then each of these persons has a claim in favor of \( w \); if, further, none are worse off in \( w \), then none have a claim in favor of \( v \); since some claims point to \( w \), and none to \( v \), it follows that \( w \) is better.) But I have also argued, in a line of reasoning summarized in the footnotes, that the claims justificatory framework supports Pigou-Dalton.\(^{55}\)

Quite apart from this specific line of justification, Pigou-Dalton has intuitive force. Suppose that we can transfer well-being from better-off Jamal to worse-off Kieran, without affecting anyone else; without losing any of the transferred well-being (a “pure” transfer); and with the well-being gap between them shrinking. Pigou-Dalton addresses this limiting case in which there is an equity “benefit” (the gap between a better-off and worse-off person shrinks) but no cost to overall well-being (the transfer is pure; overall well-being remains the same). Surely, at least in this case, equity considerations should come into play and recommend the transfer.

Minimal Pigou-Dalton is yet more intuitive. Surely there is some case in which a pure transfer from a better-off Jamal to a worse-off Kieran is a moral improvement. Yet utilitarianism denies that there is ever such a case—regardless of how much better off Jamal is than Kieran, regardless of the size of the transfer, and despite the fact that it is a pure one.

The utilitarian might defend her absence of concern for an equitable distribution of well-being (as axiomatized by Pigou-Dalton or, more weakly minimal Pigou-Dalton) by adducing the veil-of-ignorance (VOI) argument for utilitarianism, associated with John Harsanyi (1953; 1955; 1977). If we construct the outcome ranking by conceptualizing each outcome as an equiprobability lottery over the lives of the individuals therein, the upshot is a utilitarian SWF—which in turn implies rejecting Pigou-Dalton and even minimal Pigou-Dalton. In short, accepting the VOI as the basis for justifying the outcome ranking explains why Pigou-Dalton and even minimal Pigou-Dalton should be rejected, counterintuitive as that may be. But the status of the VOI as a justificatory framework is itself controversial. Is it really the case that outcome \( x \) is ethically better than outcome \( y \) iff everyone would choose \( x \) under a veil of ignorance? The dialectical situation is such that the proponent of prioritarianism can reject the VOI and instead endorse a different framework (for example, my claims-based approach) that supports Pigou-Dalton. (Adler 2012, pp. 321-337; Moreno-Ternero and Roemer 2008).

\(^{55}\) Consider a non-rank-switching Pigou-Dalton transfer. In vector \( v \), one individual (“Higher”) is at well-being level \( w_H \); a second (“Lower”) at well-being level \( w_L \), with \( w_H > w_L \). In vector \( w \), Higher’s well-being decreases by \( \Delta w > 0 \), and Lower’s well-being increases by \( \Delta w \), without switching their ranks \( (w_H - \Delta w \geq w_L + \Delta w) \), and without affecting anyone else. In this case, everyone except for Lower and Higher has a zero claim between \( v \) and \( w \). Further, it seems compelling that Lower’s claim in favor of \( w_L + \Delta w \) over \( w_L \) is stronger than Higher’s claim in favor of \( w_H - \Delta w \) over \( w_H \). (The well-being differences are the same; Lower’s well-being level is strictly lower than Higher’s at the starting point, and at least weakly lower at the end point.) Thus \( w \) is better than \( v \) in light of the pattern of claims.

This is a claims-based argument for a Pigou-Dalton axiom limited to non-rank-switching transfers. That axiom, plus Anonymity, implies the broader Pigou-Dalton axiom (for gap-diminished transfers) that we have been using in this chapter. See note 34.
A second response is that the utilitarian SWF may support equity axioms in attribute space. In particular, it is often supposed that income has diminishing marginal utility, reflected in a well-being measure that is strictly increasing and strictly concave in income. If $w(\cdot)$ is indeed strictly increasing and strictly concave in income, the utilitarian SWF will satisfy a counterpart to the Pigou-Dalton axiom as stated here, namely one framed in terms of income. But it is less clear whether non-income attributes also have diminishing marginal utility. Note further that making the ethical case for attribute transfers hinge on diminishing marginal utility—as utilitarianism does—is itself counterintuitive. Assume that I am richer than you, but am a greedy materialist who loves money so much that I gain greater well-being from a unit of income than you do. Utilitarianism has no basis for an income transfer from me to you, while prioritarianism does.

2.9 Atkinson and Kolm-Pollak SWFs

A prioritarian SWF ranks well-being vectors by means of a “transformation” function: a strictly increasing, strictly concave function, abbreviated as “$g(\cdot)$” in this chapter. Each vector $w$ is assigned a score equaling $g(w_1) + \ldots + g(w_N)$, and the vectors are ranked in the order of these scores.

But there are a multitude (an infinity!) of strictly increasing and strictly concave $g(\cdot)$ functions. Thus there are a multitude (an infinity!) of prioritarian SWFs, each defined by one such transformation function.

Two types of prioritarian SWFs are most widely discussed in the extant economic literature on prioritarianism. These are, first, Atkinson SWFs; and second, Kolm-Pollak SWFs. Each employs a quite tractable functional form for the $g(\cdot)$ function. Further, a plausible axiomatic case can be offered in defense of Atkinson SWFs. These two subfamilies of prioritarian SWFs will now be discussed in turn.

2.9.1 Atkinson SWFs

Atkinson SWFs define the $g(\cdot)$ function as follows: $g(w_i) = \frac{1}{1-\gamma} w_i^{1-\gamma}$, with $\gamma > 0$ (and $\gamma \neq 1$), and $g(w_i) = \log w_i$ in the special case of $\gamma =1$. The $\gamma$ parameter specifies the degree of priority for the worse off. The larger $\gamma$, the greater such priority. Adler (2012, pp. 378-99; 2019, pp. 154-58), discusses and defends Atkinson SWFs.

Atkinson SWFs are a subfamily of SWFs, within the broader family of prioritarian SWFs. The axiomatic argument in favor of this subfamily runs as follows. (1) To begin, let’s

---

56 Atkinson SWFs are so-called because of their role in Atkinson’s work on the measurement of inequality. (Atkinson, 1970). They are also referred to as constant-elasticity-of-substitution SWFs, and sometimes as Kolm-Atkinson-Sen SWFs because they figure in Kolm (1969) and Sen (1973). Kolm-Pollak SWFs are credited to Kolm (1969) and Pollak (1971); they are discussed in Blackorby, Bossert and Donaldson (2005, ch. 4).
assume that the prioritarian SWF is *profile-independent*. A “profile-independent” SWF is such that the functional form of the SWF is independent of our well-being measure. *Whichever* well-being measure is adopted to translate outcomes into well-being vectors—this depends on our theory of well-being—we use the very same rule to rank well-being vectors and thereby to rank outcomes via the Master Recipe. The prioritarian SWF, in particular, is profile-independent if we identify some single $g(\cdot)$ function as our transformation function, rather than a $g(\cdot)$ function that varies depending on which $w(\cdot)$ is being used.

The premise of a profile-independent SWF is often adopted in the theory literature on SWFs. Further, this premise seems to capture the spirit of welfare-consequentialism. In comparing $w$ to $v$, we should care only about the pattern of well-being in each vector, not about the various individuals’ bundles of attributes (income, health, longevity, leisure, etc.) or histories (combinations of attributes and preferences) that happen to give rise to the two patterns. The mapping *from* attribute bundles/histories *to* well-being—that is, the specific $w(\cdot)$ that is being used to convert outcomes into vectors—should be irrelevant to how we compare the patterns themselves.

(2) Second, our profile-independent prioritarian SWF should satisfy an appropriate invariance axiom, namely that if $w(\cdot)$ is swapped for an informationally equivalent well-being measure $w^*(\cdot)$, the ranking of outcomes shouldn’t change.

It is tempting to formulate this invariance requirement as follows:

**Cardinal Invariance.** Let $w^*(\cdot)$ be a common cardinal rescaling of $w(\cdot)$. Then the SWF should rank outcomes the very same way using $w^*(\cdot)$ and using $w(\cdot)$.

Note that if $w^*(\cdot)$ is a common cardinal rescaling of $w(\cdot)$, then $w^*(\cdot)$ and $w(\cdot)$ make the very same intra- and interpersonal comparisons of both well-being levels and well-being differences.

However, it can be shown that *no* profile-independent prioritarian SWF satisfies Cardinal Invariance. (Bossert and Weymark 2004, p. 1160; Weymark 2016, p. 154). This impossibility can be avoided by weakening Cardinal Invariance to Ratio Invariance.

**Ratio Invariance.** Let $w^*(\cdot)$ be a common *ratio* rescaling of $w(\cdot)$. Then the SWF should rank outcomes the very same way using $w^*(\cdot)$ and $w(\cdot)$.

Ratio invariance supposes that it is meaningful to make intra- and interpersonal comparisons of well-being levels, differences, and ratios, and that no other types of well-being comparisons are meaningful. A common ratio rescaling preserves all of the information regarding these six types

---

57 For a discussion of profile-independence, see Weymark (2016). The so-called welfarism theorem (Weymark 2016; Bossert and Weymark 2004), a key theorem in the theoretical literature on SWFs, shows that profile-independence can be derived from Pareto indifference plus an independence-of-irrelevant-alternatives axiom.

58 See above section 2.2.1.
of comparisons, and since there is no further well-being information to be had, the SWF’s outcome ranking should be invariant to such a rescaling.\(^5^9\)

To see how a common ratio rescaling preserves information regarding all six types of comparisons, suppose that there are three individuals, Amy, Barry, and Cindy. Outcome \(x\) is mapped by measure \(w(\cdot)\) onto vector \((10, 1, 5)\); \(y\) is mapped onto vector \((8, 3, 4)\); and \(z\) is mapped onto vector \((2, 2, 2)\). If we assume that intra- and interpersonal comparisons of levels, differences, and ratios are meaningful, we can “read off” various comparisons from these numbers: for example, that Amy in \(x\) is better off than Amy in \(y\) \((10 > 8)\), an intrapersonal level comparison; that Amy in \(x\) is better off than Barry in \(z\) \((10 > 2)\), an interpersonal level comparison; that the difference in well-being between Cindy in \(y\) and Cindy in \(z\) is greater than the difference in well-being between Cindy in \(x\) and Cindy in \(y\) \((4 – 2 > 5 – 4)\), an intrapersonal difference comparison; that the difference in well-being between Amy in \(x\) and Amy in \(y\) is greater than the difference in well-being between Barry in \(y\) and Barry in \(z\) \((10 – 8 > 3 – 2)\), an interpersonal difference comparison; that the well-being ratio between Amy in \(y\) and Amy in \(z\) is greater than the well-being ratio between Amy in \(x\) and Amy in \(y\) \((8/2 > 10/8)\), an intrapersonal ratio comparison; or that the well-being ratio between Amy in \(x\) and Barry in \(x\) is greater than the well-being ratio between Barry in \(x\) and Cindy in \(y\) \((10/1 > 1/4)\), an interpersonal ratio comparison.

Consider now a well-being measure \(w^*(\cdot)\), which is a common ratio rescaling of the original well-being measure. Specifically, the values according to this new well-being measure are 10 times the values assigned by the original one. So outcome \(x\) is now mapped onto the vector \((100, 10, 50)\); outcome \(y\) is mapped onto the vector \((80, 30, 40)\); and \(z\) is mapped onto the vector \((20, 20, 20)\). It is straightforward to see that all of the information about intra- and interpersonal comparisons of levels, differences, and ratios contained in the original well-being measure is preserved by this new one.\(^6^0\)

Putting together the stipulation that the prioritarian SWF be profile-independent and the axiom of Ratio Invariance, we end up with the Atkinson subfamily of SWFs. No other prioritarian SWF can satisfy both requirements. (Boadway and Bruce 1984, p. 160; Bossert and Weymark 2004, pp. 1161-62; Weymark 2016, p. 154).

In order to implement the SWF methodology using a profile-independent SWF of the Atkinson form, two steps must be taken. First, we must construct a well-being measure that contains information not only about intra- and interpersonal comparisons of levels and differences (as discussed in Section 2.5), but also intra- and interpersonal comparisons of ratios.

---

\(^5^9\) As mentioned above, \(w^*(\cdot)\) is a common cardinal rescaling of \(w(\cdot)\) if there exists \(k > 0, l\) such that \(w^*_i(x) = kw_i(x) + l\) for all \(i, x\). \(w^*(\cdot)\) is a common ratio rescaling of \(w(\cdot)\) if there exists \(k > 0\) such that \(w^*_i(x) = kw_i(x)\) for all \(i, x\).

\(^6^0\) Let \(w^*_i(x) = kw_i(x), k > 0,\) for all \(i, x\). Then, for all individuals \(i, j, k, l\) (identical or wholly or partly distinct) and outcomes \(x, y, z, z'\) (identical or wholly or partly distinct): \(w_i(x) \geq w_i(y)\) iff \(w^*_i(x) \geq w^*_i(y)\); \(w_i(x) – w_i(y) \geq w_i(z) – w_i(z')\) iff \(w^*_i(x) – w^*_i(y) \geq w^*_i(z) – w^*_i(z')\); \(w_i(x)/w_i(y) \geq w_i(z)/w_i(z')\) iff \(w^*_i(x)/w^*_i(y) \geq w^*_i(z)/w^*_i(z')\).
This can be done as follows. Assume that $w(\cdot)$ accurately reflects intra- and interpersonal level and difference information, as per our theory of well-being. Now identify a zero point, namely a particular attribute bundle or a particular history. This choice of zero point defines well-being ratios; for any two individual-outcome pairings (individual $i$ in $x$ and individual $j$ in $y$), the ratio of their well-being levels is the ratio of their well-being differences from the zero point. Rescale $w(\cdot)$ so that the zero bundle/history is assigned the number 0—that is, define a new well-being measure $w^+(\cdot)$ such that, for every individual $i$ and outcome $x$, $w^+_i(x) = w_i(x) - Z$, with $Z$ the number assigned by $w(\cdot)$ to the zero point. This new measure will contain the very same level and difference information as $w(\cdot)$ and will reflect intra- and interpersonal ratio information. (See Appendix for details.) Use this new measure $w^+(\cdot)$, in conjunction with the chosen Atkinson SWF and its uncertainty module, to rank outcomes and policies.

Second, we must specify the priority parameter $\gamma$. $\gamma$, as mentioned, determines the degree of priority for the worse off. The role of $\gamma$ can be understood by imagining leaky transfers from a better-off to a worse-off person. Imagine a better-off person at well-being level $w_H$ and a worse-off person at well-being level $w_L$. Let $K$ be the ratio between the two levels, larger divided by smaller: $K = \frac{w_H}{w_L}$. Consider a transfer in which the better-off individual loses some small amount of well-being $\Delta w$, while the worse-off one gains by some fraction of that loss, $\rho \Delta w$. If $\rho = 1$, every Atkinson SWF will approve it; this is required by the Pigou-Dalton principle. If $\rho = 0$, every Atkinson SWF will disapprove the transfer; disapproval is required by the Pareto axiom. Neither basic axiom constrains what happens in the case of leaky transfers—with $\rho$ between 0 and 1—and this is where the priority parameter $\gamma$ comes into play. The minimum acceptable fraction $\rho_{\text{min}}$ equals $(1/K)^\gamma$: it depends upon the well-being ratio and $\gamma$. This fraction decreases as the well-being ratio between the two individuals increases, and also decreases as $\gamma$ does.

Reciprocally, picking the minimum acceptable fraction for a given well-being ratio suffices to fix $\gamma$. From the equation $\rho_{\text{min}} = (1/K)^\gamma$, it follows that $\gamma = -(\log \rho_{\text{min}}/\log K)$. (See Appendix for more detail.)

2.9.2 Kolm-Pollak SWFs

Atkinson SWFs, notwithstanding their axiomatic virtues, have a significant flaw; they require well-being numbers to be non-negative or, in the case of $\gamma \geq 1$, strictly positive. The Atkinson $g(\cdot)$ function is either undefined or, if defined, not both strictly increasing and strictly concave with negative well-being numbers as inputs (or with 0 as an input for $\gamma \geq 1$).

In other words, Atkinson SWFs provide a well-defined and prioritarian ranking of well-being vectors if but only if all the numbers in all the vectors are non-negative (indeed, positive if $\gamma \geq 1$).

Further, an Atkinson SWF exhibits extreme behavior near the zero point, even with low values of $\gamma$. As an individual’s well-being level approaches the zero point, the relative weight
given to increments in her well-being, as compared to changes in the well-being of someone better off, approaches infinity.

The Kolm-Pollak family of prioritarian SWFs avoids these limitations. Well-being numbers can be positive, negative, or zero. And there is no zero point around which Kolm-Pollak SWFs exhibit extreme behavior. Further, these SWFs—like Atkinson SWFs—have a simple functional form. The transformation function is defined as follows. \( g(w_i) = -\exp(-\beta w_i) \), with \( \beta > 0 \) now the parameter that defines the degree of priority for the worse off (the counterpart to the \( \gamma \) parameter for the Atkinson SWF).

However, there is a substantial theoretical downside to Kolm-Pollak SWFs. A profile-independent Kolm-Pollak SWF will fail to satisfy the axiom of Ratio Invariance (again, Atkinson SWFs are the only prioritarian SWFs that do). In other words, a profile-independent Kolm-Pollak SWF might rank \( x \) better than \( y \) if the outcomes are represented, respectively, by the vectors \( w = (w_1, \ldots, w_N) \) and \( v = (v_1, \ldots, v_N) \); yet rank \( x \) equal to or worse than \( y \) if the outcomes are represented, respectively, by the vectors \( kw = (kw_1, \ldots, kw_N) \) and \( kv = (kv_1, \ldots, kv_N) \), \( k > 0 \)—even though representing the first outcome by \( kw \) and the second by \( kv \), rather than the first by \( w \) and the second by \( v \), is simply using a different numbering scheme to signify the very same two well-being patterns.

A closely related point is that it is not clear how to specify a profile-independent Kolm-Pollak SWF. Recall that we used the thought experiment of a leaky transfer between a better-off person, at \( w_H \) and a worse-off person, at \( w_L \), to fix the priority parameter (\( \gamma \)) for the Atkinson SWF. If we try to fix the priority parameter (\( \beta \)) for the Kolm-Pollak SWF using a parallel thought experiment, we end up with the following equation for the minimum acceptable fraction, \( \rho_{min} = e^{-\beta \min(w_H - w_L)} \). This minimum fraction now depends upon the priority parameter (\( \beta \)) and the magnitude of the well-being difference between the two persons (\( w_H - w_L \)). But it doesn’t seem meaningful to assign a magnitude to a well-being difference. What does it mean to say that the difference between Jada’s well-being and Keith’s well-being is a specific number (7, say, rather than 10)?

A workable alternative is to specify a profile-dependent Kolm-Pollak SWF. Although this approach is theoretically somewhat dubious—it goes against the spirit of welfarism—it yields a workable prioritarian alternative to the Atkinson SWF. Specifically, rather than identifying a single \( \beta \) parameter for the Kolm-Pollak SWF, we use thought experiments regarding leaky transfer in attributes from better-off to worse-off persons to identify a \( \beta \) parameter that depends upon \( w(\cdot) \). See Appendix for more details.

---

61 By contrast, the leaky-transfer thought experiment for specifying a profile-independent Atkinson SWF depends on the magnitude of the well-being ratio between the two individuals. Intuitively, it does seem meaningful to assign a magnitude to a well-being ratio. Further, there is a clear procedure for arriving at such magnitudes. If we have \( w(\cdot) \) unique up to common cardinal rescaling, and then fix a zero bundle/history, we end up with well-defined ratio magnitudes.
2.10 Prioritarian SWFs and Inequality Metrics

Inequality metrics (IMs)—such as the Gini coefficient, coefficient of variation, Atkinson inequality metric, or generalized-entropy family of metrics—are widely used in economics. (Blackorby, Bossert and Donaldson 2005, ch. 4; Chakravarty 2009; Cowell 2016; Lambert 2001.) IMs quantify the distribution of some “currency.” Much empirical work examines income inequality: in this body of work, the currency being plugged into an inequality metric is income. However, any item held by individuals can serve as the currency for inequality measurement. For example, the inequality of health, happiness, and longevity have each been studied. And, as we’ll see momentarily, well-being itself might be the currency for an IM.

Formally, an IM is a formula for assigning scores to currency vectors, quantifying the degree of inequality in each vector—a currency vector being a list of the holdings of the currency by each member of the population of interest. Standard IMs satisfy three axioms: an anonymity axiom (namely that a rearrangement of the currency among the population doesn’t change the degree of inequality); a Pigou-Dalton axiom (a pure, gap-diminishing transfer of the currency from someone with more to someone with less, leaving everyone else’s holdings unchanged, decreases the degree of inequality); and equality-as-a-lower bound (a perfectly equal distribution has the lowest inequality score). (Adler 2012, p. 115). In what follows, I use “IM” to mean a metric that satisfies these axioms.

An important theorem from the SWF literature demonstrates a linkage between SWFs and IMs. Any SWF that satisfies the Pareto, Anonymity, Pigou-Dalton and Continuity axioms has a corresponding IM. The SWF’s ranking of well-being vectors can be represented in a specific numerical form: as overall well-being discounted by the degree of inequality of well-being. Call this the Decomposition Theorem.

**Decomposition Theorem.** Let \( \succsim \) be an SWF that satisfies Pareto, Anonymity, Pigou-Dalton and Continuity. Then there exists a corresponding IM \( I^{\succsim}() \) which is such that: for any well-being vectors \( w \) and \( v \), \( w \succsim v \) iff

\[
\left( \sum_{i=1}^{N} w_i \right) \left( 1 - I^{\succsim} (w) \right) \geq \left( \sum_{i=1}^{N} v_i \right) \left( 1 - I^{\succsim} (v) \right).
\]

62 See, e.g., Adler (2015), citing studies of income, longevity and happiness inequality in examining the time trend of inequality in the US.

63 The Decomposition Theorem with the multiplicative formula as stated here is true if all well-being numbers are positive. A proof of the theorem is given in Adler 2012, pp. 119-20; it derives originally from Atkinson (1970), Kolm (1969), and Sen (1973). For a generalization of that formula that allows for negative well-being numbers, see Blackorby Bossert, and Donaldson (2005, pp. 101-02).

A different type of decomposition shows that a SWF that satisfies Pareto, Anonymity, Pigou-Dalton and Continuity can be expressed, via an additive formula, as the combination of an inequality metric and overall well-being. There exists an IM \( I^{\preceq}(\cdot) \) corresponding to the SWF \( \succsim \) such that: for any well-being vectors \( w \) and \( v \), \( w \succsim v \) iff

\[
\left( \sum_{i=1}^{N} w_i \right) - I^{\preceq} (w) \geq \left( \sum_{i=1}^{N} v_i \right) - I^{\preceq} (v) \quad .\quad \text{(Blackorby, Bossert and Donaldson, 2005, pp. 102-03; Decancq and Schokkaert, chapter 5, this volume)}.
\]

The discussion immediately below of the separability properties of \( I^{\preceq}(\cdot) \) also applies to \( I^{\succsim}(\cdot) \).
The formalism here bears close attention. There are many different IMs. \( I^p(\cdot) \) is that specific IM with a special representational property. We use well-being itself as the currency for \( I^p(\cdot) \). For each well-being vector \( w \), we calculate an inequality score, \( I^p(w) \). This score will be 0 if well-being is distributed perfectly equally in \( w \)—everyone has the same well-being number—and is bounded above by 1. So \( (1 - I^p(w)) \) is an inequality discount factor. If well-being is perfectly equal in \( w \), \( (1 - I^p(w)) = 1 \). As the distribution becomes increasingly unequal, \( (1 - I^p(w)) \) decreases and is bounded below by 0.

We now calculate a composite score for each vector: overall well-being multiplied by the inequality discount factor \( (1 - I^p(w)) \). Then the ranking of vectors according to this composite score is exactly the same as the ranking according to the SWF \( \succ \). This is what the Decomposition Theorem shows.

The Decomposition Theorem applies to prioritarian SWFs. (Note that all such SWFs satisfy the axiomatic conditions of the theorem.) It shows that prioritarianism can be expressed as the hybrid of two ethical values: overall well-being and the equality of well-being.

Further, it may well be of interest to identify the inequality metric \( I^p(.) \) corresponding to a prioritarian SWF and to use \( I^p(.) \) alone to rank vectors and policies. This shows how vectors and policies compare with respect to the ethical value of equalizing well-being, with the degree of attainment of that value quantified by the inequality metric implicit in the SWF. But there’s a difficulty: doing this sacrifices the pragmatic advantages of Separability.

Recall that prioritarian SWFs satisfy the Separability axiom with respect to the vector ranking. Moreover, every prioritarian SWF has an uncertainty module (specifically EPP\(^{64}\)) that satisfies Policy Separability. The decisionmaker who adopts a prioritarian SWF and ranks policies with the EPP uncertainty module can “bracket” the well-being of individuals in the population of ethical concern who are sure-to-be-unaffected (e.g., prior generations). The decisionmaker can focus her modelling and information-gathering efforts on the well-being only of potentially affected individuals (individuals who are not sure-to-be-unaffected, e.g., the present and future generations). Because her uncertainty module is policy-separable, she need not assign probabilities to the well-being of those sure-to-be-unaffected.\(^{65}\)

However, no IM satisfies the Separability axiom. Even if the SWF \( \succ \) is prioritarian and hence satisfies Separability, its corresponding IM \( I^p(.) \) will not. Table 2.11 illustrates why no IM can satisfy Separability.

---

\(^{64}\) EAP also satisfies Policy Separability, but is problematic because it fails Dominance. See above Section 2.7.

\(^{65}\) The advantages of Policy Separability were discussed in Section 2.4.
Table 2.11: Inequality metrics (IMs) violate Separability

<table>
<thead>
<tr>
<th></th>
<th>w</th>
<th></th>
<th>v</th>
<th></th>
<th>w*</th>
<th></th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris</td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td>10</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Joel</td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td>10</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Karl</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Laura</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

Explanation: Note that well-being is perfectly equally distributed in w but not v and in v* but not w*. Thus any IM should assign a lower degree of inequality to w than v and to v* than w*. However, this violates Separability. The w/v pair is the same as the w*/v* pair except for the fixed well-being levels of Karl and Laura—each fixed at 10 in the first case and 20 in the second.

Indeed, it is quite intuitive that the \( I^>(\cdot) \) ranking is not separable. \( I^>(w) < I^>(v) \) indicates that well-being is more equally distributed in vector w than vector v. But in order to compare w and v with respect to inequality, we can’t focus our attention just on the subset of affected individuals, those who are better off in w than v or vice versa. The degree of well-being inequality in w depends on the well-being levels of everyone in the population—including those who are equally well off in w and v. Similarly, the degree of well-being inequality in v depends on the well-being levels of everyone in the population—including those who are equally well off in w and v.

In order to use an IM to evaluate policies, we need an uncertainty module for the IM. We need a procedure for ranking policies understood as state-by-state arrays of well-being vectors, just as we do for an SWF. But because no IM satisfies Separability, no IM has an uncertainty module that satisfies Policy Separability. In considering how policies fare with respect to the inequality of well-being, the decisionmaker cannot “bracket” the well-being of those sure-to-be-affected (e.g., prior generations). The deliberational advantages that flow from Separability are not available when it comes to inequality assessment, even using the inequality metric that corresponds to a prioritarian and hence separable SWF.

2.11 Variable Population

The fixed-and-finite population setup, adopted until now, assumes a set of outcomes \( O = \{x, y, \ldots \} \) and a finite set of individuals \( I = \{1, \ldots, N\} \), each of whom exists in every outcome in \( O \). But it’s surely the case, for any individual, that this person might not have come into being. No person exists in every possible world. Of course, outcomes are cognitively tractable models of possible worlds, and \( O \) is a model of the set of all possible worlds. In many contexts, it may be adequate to employ the SWF framework with a fixed population. But in some contexts, this is not adequate. In particular, certain policies (for example, those concerning climate change) can be expected to significantly alter the number and identities of future individuals. If the SWF framework is to provide ethical guidance regarding such alterations, the fixed-population assumption needs to be relaxed.
In what follows, I continue to assume a finite number of individuals who exist in a given outcome, but relax the assumption that the population is fixed. As shorthand, I’ll refer to this as the “variable-population” setup and to the case of a fixed-and-finite population as the “fixed-population” setup.66

In the variable-population setup, each outcome in \( \mathbf{O} = \{x, y, \ldots \} \) has a finite number of existing individuals, but the number and identities of individuals may vary between outcomes. \( \mathbf{I} = \{1, 2, \ldots \} \) is now everyone who exists in at least one outcome in \( \mathbf{O} \). \( \mathbf{I} \) may be finite or infinite in size.67

A well-being vector has a slot for each person in \( \mathbf{I} \). If \( \mathbf{I} \) is finite, then these are finite-length vectors; otherwise they are infinite vectors. Let’s use “\( \Omega \)” as a symbol to indicate an individual’s nonexistence. Each slot in a well-being vector has as its entry either a well-being number or \( \Omega \). For example, if there are five individuals in \( \mathbf{I} \), then each well-being vector has five slots, and one such vector is as follows: \((27, \Omega, 15, 6, \Omega)\). This indicates that individual 1 is at well-being level 27; individual 2 does not exist; individual 3 is at level 15; individual 4 is at level 6; individual 5 does not exist. Or, if there are an infinite number of individuals in \( \mathbf{I} \), then each well-being vector has an infinite number of slots and might be as follows: \((\Omega, 67, 8, \Omega, \ldots)\), indicating that individual 1 does not exist; individual 2 is at well-being level 67; individual 3 is at well-being level 8; individual 4 does not exist; and so on.

An SWF is a rule for ranking well-being vectors. As in the fixed-population context, each outcome is mapped by the well-being measure onto a well-being vector, and the ranking of outcomes is according to the ranking by the SWF of their corresponding vectors.

We can simplify a bit. It’s very plausible to adopt a generalized anonymity axiom which implies that any vector with \( M \) existing individuals at various well-being levels is equally good as a vector in which the first \( M \) individuals are at those well-being levels and no one else exists. For example, the vector \((27, \Omega, 15, 6, \Omega)\) is equally good as \((27, 15, 6, \Omega, \Omega)\). Thus the ranking of \((27, \Omega, 15, 6, \Omega)\) versus \((\Omega, 9, \Omega, 12, \Omega)\) should be the same as the ranking of \((27, 15, 6, \Omega, \Omega)\) versus \((9, 12, \Omega, \Omega, \Omega)\).

We can therefore express the SWF’s ranking of vectors that may have “\( \Omega \)” entries in terms of a ranking of “reduced” vectors that lack such entries: vectors such as \((27, 15, 6, \Omega, \Omega)\) \((9, 12, \Omega, \Omega, \Omega)\). The ranking of two such “reduced” vectors means that any pair of vectors of the same length with those numerical entries and any number of “\( \Omega \)” entries is ranked accordingly. That is to say: the SWF’s ranking of the vector \((27, 15, 6)\) above \((9, 12)\) means: it prefers \((27, 15, 6)\) to \((9, 12, \Omega)\); it prefers \((27, 15, 6, \Omega)\) to \((9, 12, \Omega, \Omega)\); it prefers \((27, 15, 6, \Omega, \Omega)\) to \((9, 12, \Omega, \Omega, \Omega)\). The SWF’s rankings of the vector \((3, 3)\) above \((2, 5)\) means: it prefers \((3, 3)\) above \((2, 5)\).

---

66 See above note 29, noting the possibility of a countably infinite number of individuals who exist in a given outcome. That possibility is not addressed here.

3) to (2, 5); (3, 3, Ω) to (2, 5, Ω); and so forth. In what follows, I express variable-population SWFs in these terms.

In the fixed-population case, of course, the vectors being compared all have the same number of entries. Each SWF for the fixed-population case has a family of variable-number extensions. All such extensions converge in their ranking of same-number vectors, but differ in how they compare vectors with different numbers of entries.

For example, the utilitarian SWF compares \( \mathbf{w} = (27, 15, 6) \) to \( \mathbf{v} = (14, 16, 15) \) by assigning each a score equaling the sum of its entries. \( \mathbf{w} \) is assigned a score equaling 27+15+6 = 48, \( \mathbf{v} \) a score equaling 14+16+15 = 45, and so \( \mathbf{w} \) is ranked better than \( \mathbf{v} \). Each different-number extension of the utilitarian SWF will (a) rank \( \mathbf{w} \) better than \( \mathbf{v} \), and (b) tell us how to make different-number comparisons, e.g., how to compare (27, 15, 6) to (9, 12).

Similarly, a given prioritarian SWF is identified by some strictly increasing and strictly concave transformation function \( g(\cdot) \). \( \mathbf{w} \) is assigned a score equaling \( g(27)+g(15)+g(6) \), \( \mathbf{v} \) a score equaling \( g(14) + g(16) + g(15) \), and the two are ranked according to these scores. Each different-number extension of this prioritarian SWF will (a) preserve this ranking of \( \mathbf{w} \) versus \( \mathbf{v} \) and (b) tell us how to make different-number comparisons, e.g., how to compare (27, 15, 6) to (9, 12).

A variety of different-number extensions of the utilitarian SWF have been described in the literature. The three possibilities that have prompted the most discussion are “total” utilitarianism, “average” utilitarianism, and “critical-level” utilitarianism. The scoring system used by each is as follows. \( w_{\text{neutral}} \) is the well-being level of the “neutral” life: a life equally good for well-being as non-existence. \( w_{\text{critical}} \) is the well-being level of the “critical-level” life. \( N(\mathbf{w}) \) is the number of numerical entries. Total utilitarianism: \( \sum (w_i - w_{\text{neutral}}) \). Average utilitarianism: \( \frac{\sum w_i}{N(\mathbf{w})} \). Critical-level utilitarianism: \( \sum (w_i - w_{\text{critical}}) \). For simplicity, in the illustrations that follow, I’ll assume that \( w_{\text{neutral}} = 0.58 \).

There is a vast philosophical literature regarding the pros and cons of total, average, and critical-level utilitarianism and other variable-population extensions of utilitarianism.\(^{69}\) Suffice it to say that utilitarians are far from consensus about how the ethical goal of maximizing overall well-being should be generalized to outcomes with different numbers of individuals. The reason, in a nutshell, is that no variable-population SWF satisfies all the axioms that seem initially quite

\(^{68}\) \( w_{\text{neutral}} \) is often set to zero in the literature on population ethics, but this is not necessary. Indeed, setting \( w_{\text{neutral}} = 0 \) is problematic if the SWF is an Atkinson prioritarian SWF rather than a utilitarian SWF. As explained in Section 2.9.2, the Atkinson \( g(\cdot) \) function cannot take negative well-being numbers as inputs. Setting \( w_{\text{neutral}} = 0 \) with an Atkinson SWF precludes well-being vectors with lives worse than neutral. Instead, with an Atkinson SWF, we should set \( w_{\text{neutral}} > 0 \), which allows for lives worse than neutral but with well-being values above 0. (Adler and Treich 2015).

However, the examples in the text are illustrating the features of variable-population utilitarianism, and so I simplify by setting \( w_{\text{neutral}} = 0 \).

\(^{69}\) See Broome (2016), Greaves (2017) for overviews and Arrhenius, Bykvist, Campbell and Finneron-Burns (forthcoming) for an authoritative handbook.
plausible. Total utilitarianism runs afoul of the so-called “Repugnant Conclusion.” For (a) any level of well-being \( w^h \) however high, (b) any vector \( w \) with every individual at level \( w^h \), and (c) any level of well-being \( w^f \) arbitrarily close to neutral, \( w^f > w^{neutral} \), there is a vector \( v \) with a sufficiently large number of individuals all at \( w^f \) which is better than \( w \). For example, let \( w^h = 1000 \) and \( w^f = 1 \). Then a well-being vector with \( M \) individuals at level 1000 is worse than any well-being vector with a much larger population (a population with 1000\( M + k \) individuals, \( k \geq 1 \)) at level 1.

Average utilitarianism avoids the Repugnant Conclusion, but has other difficulties. It is not separable in making different-number comparisons: (4, 6, 11) is preferred to (4, 6), but (20, 6, 11) is dispreferred to (20, 6). Average utilitarianism also violates the “Negative Expansion” principle: adding an individual who is worse off than \( w^{neutral} \) should not be an improvement. According to average utilitarianism, \((-10, -8, -3)\) is better than \((-10, -8)\). Finally, and reciprocally, it violates the “Mere Addition” axiom: adding an individual with a life better than \( w^{neutral} \) should not be a worsening. \((15, 25, 10)\) is worse than \((15, 25)\).

If \( w^{critical} > w^{neutral} \), critical-level utilitarianism avoids the Repugnant Conclusion, satisfies the Negative Expansion principle, and is separable in making different-number comparisons. However it violates Mere Addition and also violates “Priority for Lives Worth Living”: It shouldn’t be the case that a vector with everyone above the neutral level is worse than one with everyone below the neutral level. For example, assume that \( w^{critical} = 5 \). Then \((2, 4, 3, 1, 2)\) is assigned a critical level utilitarian score of \(-13\), which is less than the score of \((-1, -1)\).

Prioritarian SWFs have different-number extensions analogous to those of utilitarianism. In particular, “total prioritarianism,” “average prioritarianism” and “critical-level prioritarianism” can all be described, as follows. **Total prioritarianism**: \( \sum (g(w_i) - g(w^{neutral})) \). **Average prioritarianism**: \( \frac{\sum g(w_i)}{N(w)} \). **Critical-level prioritarianism**: \( \sum (g(w_i) - g(w^{critical})) \). The axiomatic features of these approaches with respect to the Repugnant Conclusion, Different Number Separability, the Negative Expansion principle, Mere Addition and Priority for Lives Worth Living are the same as those of their utilitarian analogues. See Table 2.12, summarizing the axiomatic properties of total, average, and critical-level utilitarianism and prioritarianism.

---

70 See Adler (2019, pp. 244-48), discussing the axiomatic features of total, average, and critical-level utilitarianism described in the next several paragraphs.

71 Different-Number Separability requires that the ranking of two well-being vectors with different numbers of entries be independent of the well-being levels of individuals with the same well-being levels in both entries. This is what Blackorby, Bossert and Donaldson (2005, p. 159) term “Utility Independence.”
Table 2.12: Total, average, and critical-level utilitarianism and prioritarianism

<table>
<thead>
<tr>
<th></th>
<th>Total utilitarianism and prioritarianism</th>
<th>Average utilitarianism and prioritarianism</th>
<th>Critical-level utilitarianism and prioritarianism (with $w_{\text{critical}} &gt; w_{\text{neutral}}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoids Repugnant Conclusion</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Different-Number Separability</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Negative Expansion</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Mere Addition</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Priority for Lives Worth Living</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

I’ve said nothing thus far about policies in the variable-population case. The treatment, here, is the same as for the fixed-population case. There is a set $P$ of policies {$P, P^*, \ldots$}. We continue to use the state-of-nature setup: a policy is a state-by-state array of outcomes and thus a state-by-state array of well-being vectors (replacing each outcome with its vector). Any given SWF for the fixed-population case has a whole family of different-number extensions, and each of these in turn has multiple uncertainty modules. For example, the total-utilitarian SWF has multiple modules, as does average and critical-level utilitarianism, and the same is true for total, average, and critical-level prioritarianism.

2.12 Conclusion

This chapter has explicated the theory of prioritarian SWFs. The SWF framework—the most rigorous methodology for implementing welfare-consequentialism—converts outcomes (models of possible worlds) into well-being vectors, and ranks the vectors using some rule (the SWF). Government policies are conceptualized as probability distributions over outcomes; and a given SWF is paired with an uncertainty module, for ranking policies thus conceptualized.

Prioritarianism is a family of SWFs. Each such SWF is defined by a strictly increasing and strictly concave transformation function, and ranks vectors according to the sum of well-being numbers plugged into this transformation function. All such SWFs satisfy the axioms of Pareto, Anonymity, Separability, Continuity, and Pigou-Dalton; and the converse is also true, namely that only prioritarian SWFs satisfy these axioms. The main uncertainty modules for a prioritarian SWF are ex post prioritarianism (EPP); ex ante prioritarianism (EAP); and expected equally-distributed-equivalent prioritarianism (EEDEP). These were compared with respect to uncertainty axioms, namely: ex ante Pareto, Dominance, and Policy Separability.

The chapter also reviewed two important subfamilies of prioritarian SWFs: Atkinson SWF and Kolm-Pollak SWFs. These both employ specific and quite tractable formulas for the prioritarian transformation function. Atkinson and/or Kolm-Pollak SWFs are the versions of
Prioritarianism generally employed in the *Prioritarianism in Practice* volume and in previous applied work.

It is sometimes argued that prioritarianism is no different from utilitarianism. What animates the argument is a simple mathematical observation. Prioritarianism assigns a given outcome $x$ the score $g(w_1(x)) + \ldots + g(w_N(x))$, with $w(\cdot)$ the well-being measure and $g(\cdot)$ the prioritarian transformation function. If we now define a new well-being measure $w^*(\cdot)$ as follows—$w^*_i(x) = g(w_i(x))$—the utilitarian outcome ranking using $w^*(\cdot)$ is identical to the prioritarian outcome ranking using $g(\cdot)$ and $w(\cdot)$.

The mathematical observation is correct, but to leap from that observation to the conclusion that prioritarianism is just utilitarianism under a different name is problematic. Relative to any given well-being measure, the prioritarian and utilitarian outcome and policy rankings will generally be different. Moreover, the two measures immediately above, $w(\cdot)$ and $w^*(\cdot)$, are not the same. They imply distinct intra- and interpersonal comparisons of well-being differences; and a theory of well-being, insofar as it allows for such comparisons, will see the two well-being measures as distinct.²²

The chapter not only explicated the theory of prioritarian SWFs, but also set forth a normative defense of prioritarianism. The key normative failing of utilitarianism is its insensitivity to the distribution of well-being. The utilitarian SWF fails to satisfy the Pigou-Dalton axiom, which stipulates that a gap-diminishing pure transfer of well-being from a better-off to a worse-off person is an ethical improvement. Indeed, the utilitarian SWF is such that a transfer of well-being from a better- to a worse-off person that leaves unchanged the sum total of well-being is *always* a matter of ethical indifference—regardless of how much better off the better-off one.

By contrast, prioritarianism satisfies Pigou-Dalton. To be sure, that is also true of other SWFs—most prominently the leximin SWF, rank-weighted SWFs, and sufficientist SWFs. But unlike these other equity-regarding, non-utilitarian SWFs, prioritarian SWFs are especially tractable—an important consideration for a policy-analysis tool—in virtue of also satisfying the Separability and Continuity axioms. Pareto, Anonymity Separability, and Continuity, taken together, characterize the *generalized utilitarian* class of SWFs: all SWFs that sum a strictly increasing and continuous function of well-being. In the case of utilitarianism, this function is the identity function (utilitarianism simply adds up untransformed well-being). If we require, however, that a generalized-utilitarian SWF also be fully equity-regarding—that it satisfy Pigou-Dalton—we end up with prioritarianism.

²² Section 2.5 argued that well-being difference comparisons, not merely level comparisons, should be seen as admissible.
2.A Appendix

The SWF framework has the following components. \( O = \{ x, y, \ldots \} \) is a set of outcomes. \( I = \{ 1, 2, \ldots, N \} \) is a finite population of individuals, each of whom exists in all of the outcomes. \( N \geq 3 \). \( w(\cdot) \) is a well-being measure, with \( w_i(x) \) the well-being number of individual \( i \) in outcome \( x \). \( x \) corresponds to the \( N \)-entry well-being vector \((w_1(x), w_2(x), \ldots, w_N(x))\). Well-being vectors are also denoted with bold-faced lower case letters such as \( w \) or \( v \). \( W \) is the set of all \( N \)-entry well-being vectors.

An SWF, \( \succ \), is an ordering of \( W \). A quasiordering is a transitive, reflexive, binary relation and an ordering is a complete quasiordering. Transitive: For all \( w, v, u \), if \( w \succ v \) and \( v \succ u \) then \( w \succ u \). Reflexive: For all \( w, w \succ w \). Complete: For all \( w, v \), either \( w \succ v \) or \( v \succ w \) or both. \( w \succ v \) should be read as “\( w \) is at least as good as \( v \).” From this at-least-as-good relation, we generate an equally-good-as-relation (\( \sim \)) and better-than relation (\( \succ \)) as follows: \( w \sim v \) iff \( w \succ v \) and \( v \succ w \). \( w \succ v \) iff \( w \succ v \) and not \( v \succ w \). It follows that: \( w \succ v \) iff \( w \succ v \) or \( w \sim v \). (One item at least as good as a second iff equally good or better than the second.)

An SWF ranks outcomes using the Master Recipe: \( x \) at least as good as \( y \) iff \( (w_1(x), w_2(x), \ldots, w_N(x)) \succeq (w_1(y), w_2(y), \ldots, w_N(y)) \).

\( P = \{ P, P^*, \ldots \} \) is the set of policies, each understood as a state-by-state array of outcomes. A given SWF has different possible uncertainty modules. An uncertainty module specifies an ordering of \( P \), denoted here as \( \succeq^P \).

In what follows, I first present formulas for the major SWFs and the axioms that characterize them; then discuss the policy ranking and uncertainty modules; and finally review the Atkinson and Kolm-Pollak functional forms for prioritarianism. Sources discussing the SWFs, axioms, and uncertainty are provided in sections 2.2 and 2.3 of the main text; and discussing the Atkinson and Kolm-Pollak functional forms in section 2.9.

2.A.1 SWFs: Formulas and Axioms

Major types of SWFs are as follows. (These formulas are given in Table 2.1 and, for convenience, are repeated here.)

---

73 Or all in some orthant of \( N \)-dimensional space. The possible need to restrict well-being vectors to an orthant is illustrated by Atkinson SWFs, which require well-being numbers to be non-negative (or positive if \( \gamma \geq 1 \)).
The Utilitarian SWF:  \( w \succeq v \) iff \( \sum_{i=1}^{N} w_i \geq \sum_{i=1}^{N} v_i \).

The Prioritarian Family of SWFs:  \( w \succeq v \) iff \( \sum_{i=1}^{N} g(w_i) \geq \sum_{i=1}^{N} g(v_i) \), with \( g(\cdot) \) a strictly increasing and strictly concave function.\(^{74}\)

In presenting the next two types of SWF, I’ll need some additional symbolism. For a given well-being vector \( w \), let \( \hat{w} \) be a vector listing the elements of \( w \) from smallest to largest.

The Leximin SWF:  For any two vectors \( w, v \): (1) \( w \sim v \) iff \( w \) is a permutation of \( v \); and (2) \( w \succ v \) iff there is \( j \leq N \) such that \( \hat{w}_i = \hat{v}_i \) for all \( i < j \) and \( \hat{w}_j > \hat{v}_j \).

The rank-weighted family of SWFs:  Each SWF within this family is defined by a list of \( N \) positive and strictly decreasing weights: \( k_1, k_2, \ldots, k_N \) such that \( k_1 > k_2 > \ldots > k_N > 0 \).

\( w \succeq v \) iff \( \sum_{i=1}^{N} k_i \hat{w}_i \geq \sum_{i=1}^{N} k_i \hat{v}_i \).

The sufficientist family of SWFs:  Each SWF within this family is defined by some threshold level of well-being \( w_{\text{thresh}} \), and some strictly increasing and strictly concave function \( g(\cdot) \). A given well-being vector \( w \) has two associated vectors: \( \overline{w} \), the elements of \( w \) truncated above at \( w_{\text{thresh}} \); and \( \underline{w} \), the elements of \( w \) truncated below at \( w_{\text{thresh}} \). (That is, \( \overline{w}_i = \min \{ w_i, w_{\text{thresh}} \} \); and \( \underline{w}_i = \max \{ w_i, w_{\text{thresh}} \} \). For any two vectors \( w, v \): (1) \( w > v \) iff either (a) \( g(\overline{w}_i) + \ldots + g(\overline{w}_N) > g(\underline{v}_i) + \ldots + g(\underline{v}_N) \) or (b) \( g(\overline{w}_1) + \ldots + g(\overline{w}_N) = g(\underline{v}_1) + \ldots + g(\underline{v}_N) \) and \( \underline{w}_1 + \ldots + \underline{w}_N > \underline{v}_1 + \ldots + \underline{v}_N \); and (2) \( w \sim v \) iff \( g(\overline{w}_1) + \ldots + g(\overline{w}_N) = g(\underline{v}_1) + \ldots + g(\underline{v}_N) \) and \( \underline{w}_1 + \ldots + \underline{w}_N = \underline{v}_1 + \ldots + \underline{v}_N \).

The utilitarian SWF and prioritarian SWFs fall within a broader class of SWFs, namely generalized-utilitarian SWFs:

The generalized-utilitarian class of SWFs:  \( w \succeq v \) iff \( \sum_{i=1}^{N} f(w_i) \geq \sum_{i=1}^{N} f(v_i) \), with \( f(\cdot) \) strictly increasing and continuous.

At this juncture, it’s worth introducing the “equally distributed equivalent” (EDE) representation of a prioritarian SWF. Rank-weighted SWFs and generalized-utilitarian SWFs are “score-based”: they rank vectors according to the values assigned to vectors by a real-valued function. For any such SWF, the ranking of vectors according to a given real-valued function, and according to any strictly increasing transformation of that function, are the same. In particular, consider a prioritarian SWF as defined immediately above. Let \( H(\cdot) \) be any strictly increasing function. Then the ranking of vectors using the formula \( \sum_{i=1}^{N} g(w_i) \) is the same as that

\(^{74}\) \( g(\cdot) \) is also continuous. If the domain of \( g(\cdot) \) is an open interval, this is automatic, since concave functions are continuous on open intervals.
using the formula $H\left(\sum_{i=1}^{N} g(w_i)\right)$:

$$\sum_{i=1}^{N} g(w_i) \geq \sum_{i=1}^{N} g(v_i) \iff H\left(\sum_{i=1}^{N} g(w_i)\right) \geq H\left(\sum_{i=1}^{N} g(v_i)\right).$$

For a given prioritarian transformation function (a strictly increasing and strictly concave function $g(\cdot)$), the equally distributed equivalent for a given well-being vector $w$, $w^{EDE}(w)$, is the well-being level such that a vector with all $N$ individuals at that level is equally good as $w$. 

$$w^{EDE}(w) = g^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} g(w_i)\right).$$

Note now that $w^{EDE}(w) = H\left(\sum_{i=1}^{N} g(w_i)\right)$, with $H(m) = g^{-1}(m/N)$, and $H(\cdot)$ strictly increasing.

In short, the straight prioritarian formula $\sum_{i=1}^{N} g(w_i)$ and the equally-distributed-equivalent formula $g^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} g(w_i)\right)$ yield exactly the same rankings of well-being vectors.

The choice between the two becomes significant in constructing an uncertainty module: the ranking of policies according to the expected value of the first formula (the EPP module) is not generally the same as the ranking of policies according to the expected value of the second formula (the EEDEP module). See Section 2.7; 2.A.2. The choice is also significant if one wishes not merely to rank well-being vectors, but to measure their social-welfare value on a supra-ordinal scale (e.g., a ratio or translation scale). (Decancq and Schokkaert, chapter 5, this volume).

The most important axioms for characterizing SWFs, as discussed in the text, are as follows:

**Pareto Indifference:** If $w_i = v_i$ for all $i$, then $w \sim v$.\(^{75}\)

**Strong Pareto:** If $v_i \geq w_i$ for all $i$, and $v_j > w_j$ for at least one $j$, then $v > w$.

**Anonymity:** Let $\pi(\cdot)$ be a permutation mapping (a one-to-one, onto mapping) from the set of individuals $I$ to $I$. $\pi(\cdot)$ associates each individual $i$ with one and only one individual $\pi(i)$. If $w_i = v_{\pi(i)}$ for all $i$, $w \sim v$.

**Pigou-Dalton:** Let $w$ and $v$ be such that: $w_i > w_j$ for some $i$ and $j$; $v_i = w_i - \Delta w$, $v_j = w_j + \Delta w$, with $\Delta w > 0$; $|w_i - w_j| \geq |v_i - v_j|$; and for all $k \neq i, j$, $w_k = v_k$. Then $v > w$.

**Separability:** Let $M$ be any subset of $I$, and let $M^+ = I \setminus M$ (all individuals not in M). Assume $w$, $v$, $w^*$, $v^*$ are as follows. For all $i \in M$, $w_i = v_i$ and $w_i^* = v_i^*$. For all $j \in M^*$, $w_j = w_j^*$ and $v_j = v_j^*$. Then $w \succeq v$ iff $w^* \succeq v^*$.

\(^{75}\) This is automatic, since $\succeq$ is just a ranking of well-being vectors, and can’t differentiate between vectors with the same numbers in each of the $N$ slots.
Continuity: Recall that \( W \) is the set of all well-being vectors. For any \( w \), the sets \( \{ v \in W : v > w \} \) and \( \{ v \in W : w > v \} \) are open sets. Equivalently, the sets \( \{ v \in W : v \geq w \} \) and \( \{ v \in W : w \geq v \} \) are closed sets.

In comparing utilitarianism and prioritarianism (Section 2.8), the chapter introduces another axiom, Minimal Pigou-Dalton.

Minimal Pigou-Dalton: (1) There exist \( w \) and \( v \) such that \( v > w \) and: \( w_i > w_j \) for some \( i \) and \( j \); \( v_i = w_i - \Delta w \), \( v_j = w_j + \Delta w \), with \( \Delta w > 0 \); \( v_i - v_j \); and for all \( k \neq i, j \), \( w_k = v_k \). (2) Let \( w \) and \( v \) be such that: \( w_i \geq w_j \) for some \( i \) and \( j \); \( v_i = w_i + \Delta w \), \( v_j = w_j - \Delta w \), with \( \Delta w > 0 \); and for all \( k \neq i, j \), \( w_k = v_k \). Then \( w \geq v \).

2.A.2 Uncertainty and the Policy Ranking

Uncertainty is captured by a set \( S = \{ s, s^*, \ldots \} \) of states of nature. \( S \) is finite. \( \pi(s) \) is the probability of state \( s \). \( \sum_{s \in S} \pi(s) = 1 \). Zero-probability states can be ignored, and so I assume that \( \pi(s) > 0 \) for all \( s \). There is a set \( P \) of policies, \( P = \{ P, P^*, \ldots \} \). A policy associates each state with some outcome; this is the outcome that would occur, were that policy to be chosen and that state to be the actual state. Let \( x_{P,s} \) denote the outcome in state \( s \) of policy \( P \).

An uncertainty module for a given SWF \( \succeq \) specifies an ordering \( \succeq^P \) of the set of policies \( P \). The module is a formula for arriving at \( \succeq^P \), for any set of outcomes \( O \), set of states \( S \) and state probabilities, set of policies \( P \), and well-being measure \( w(\cdot) \). I assume that this formula takes account of the state probabilities and of the well-being vectors in each state (as per \( w(\cdot) \)), but is independent of the description of the states and of which specific outcome gives rise to the well-being vector in a given state.

All the modules for a given SWF converge in their ranking of “degenerate” policies—a degenerate policy being one that yields the same well-being vector in each state. This can be captured in a requirement of “Basic Consistency” which every module must satisfy.

Basic Consistency: Assume that (1) there exists some \( w \), such that \( (w_1(x_{P,s}), \ldots, w_N(x_{P,s})) = w \) for every state \( s \), and (2) there exists some \( w^* \) such that \( (w_1(x_{P^*,s}), \ldots, w_N(x_{P^*,s})) = w^* \) for every state \( s \). Then \( P \succeq^P P^* \) iff \( w \succeq w^* \).

The leading modules for the prioritarian SWF, as discussed in the text, are “ex post prioritarianism” (EPP); “ex ante prioritarianism” (EAP); and expected equally-distributed-equivalent prioritarianism (EEDEP). Their formulas are as follows. Throughout, \( g(\cdot) \) is a strictly increasing and strictly concave function. As shorthand, I’ll use \( w_i(P) \) to denote \( i \)’s expected well-being with policy \( P \). \( w_i(P) = \sum_{s \in S} \pi(s)w_i(x_{P,s}) \).
EPP: \( P \succeq^P P^* \iff \sum_{s \in S} \pi(s) \sum_{i=1}^{N} g(w_i(x_{p,s})) \geq \sum_{i=1}^{N} \pi(s) g(w_i(x_{p,s})) \). Note that this formula is equivalent to summing individuals’ expected transformed well-being.

\[
\sum_{s \in S} \pi(s) \sum_{i=1}^{N} g(w_i(x_{p,s})) = \sum_{i=1}^{N} \sum_{s \in S} \pi(s) g(w_i(x_{p,s}))
\]

EAP: \( P \succeq^P P^* \iff \sum_{i=1}^{N} g(w_i(P)) \geq \sum_{i=1}^{N} g(w_i(P^*)) \).

EEDEP: \( P \succeq^P P^* \iff \sum_{s \in S} \pi(s) g^{-1} \left( \sum_{i=1}^{N} g(w_i(x_{p,s})) / N \right) \geq \sum_{i=1}^{N} \pi(s) g^{-1} \left( \sum_{i=1}^{N} g(w_i(x_{p,s})) / N \right) \).

The module for the utilitarian SWF, discussed in the text, is UUU.

UUU: \( P \succeq^P P^* \iff \sum_{s \in S} \pi(s) \sum_{i=1}^{N} w_i(x_{p,s}) \geq \sum_{i=1}^{N} w_i(x_{p,s}) \). Note that this formula is equivalent to summing individuals’ expected well-being.

\[
\sum_{s \in S} \pi(s) \sum_{i=1}^{N} w_i(x_{p,s}) = \sum_{i=1}^{N} \sum_{s \in S} \pi(s) w_i(x_{p,s}) = \sum_{i=1}^{N} w_i(P)
\]

The uncertainty axioms that orient discussion in the text are the ex ante Pareto axiom (in two parts, Indifference and Strong); Dominance; and Policy Separability.

Ex Ante Pareto Indifference: If \( w_i(P) = w_i(P^*) \) for all \( i \), then \( P \sim^P P^* \).

Ex Ante Strong Pareto: If \( w_i(P) \geq w_i(P^*) \) for all \( i \), and \( w_j(P) > w_j(P^*) \) for at least one \( j \), then \( P \succ^P P^* \).

Dominance: Assume that, for all \( s \in S, (w_1(x_{p,s}), \ldots, w_N(x_{p,s})) \succ (w_1(x_{p^*,s}), \ldots, w_N(x_{p^*,s})). \) Then \( P \succ^P P^* \).

Policy Separability: Let \( M \) be any subset of \( I \) (the set of individuals), and let \( M^+ = I \setminus M \) (all individuals not in \( M \)). Assume policies \( P, P^+, P'^+ \) are as follows. For each \( i \in M, w_i(x_{p,s}) = w_i(x_{p^*,s}) \) and \( w_i(x_{p^+,s}) = w_i(x_{p^+,s}) \) for every state \( s \). For each \( j \in M^+, w_j(x_{p,s}) = w_j(x_{p^+,s}) \) and \( w_j(x_{p^+,s}) = w_j(x_{p^+,s}) \) for every state \( s \). Then \( P \succeq^P P^* \iff P^+ \succeq^P P'^+ \).

2.A.3 Functional Forms for Prioritarianism: Atkinson and Kolm Pollak

The analysis that follows is formulated in term of a preference-based view of well-being. An individual’s well-being depends upon her history \( h \), with a history a combination of a bundle of non-preference attributes \( a \), and a preference structure \( R. \ h = (a, R). \) (As explained in Adler
This Section will be switching back and forth between a well-being measure that assigns well-being numbers to individuals in outcomes, and a well-being measure that assigns well-being numbers to histories. In order to avoid confusion, I’ll continue to use \( w(\cdot) \) to denote the former, and will use \( \theta(\cdot) \) to denote the latter. \( w_i(x) = \theta(h_i(x)) = \theta(a_i(x), R_i(x)) \).

I assume, throughout, that \( \theta(\cdot) \) is unique up to a positive affine transformation. If \( \theta(\cdot) \) is unique up to a positive affine transformation, then \( w(\cdot) \) is unique up to a common cardinal rescaling and represents both intra- and interpersonal comparisons of both well-being levels and differences.  

The analysis here translates in a straightforward way to non-preference-based views of well-being. This case is simpler, since the \( R \) can be dropped. \( w_i(x) = \theta(a_i(x)) \).

2.A.3.1 Atkinson SWF (Profile-Independent)

As discussed in the text, the theoretical literature on SWFs generally assumes that the SWF is profile-independent. \( \succeq \) is a single ranking of well-being vectors, which is independent of \( w(\cdot) \). Assume that \( \succeq \) ranks well-being vectors \( w, v, ww, \) and \( vv \) a certain way. Imagine that we are using \( w(\cdot) \) to assign well-being numbers to individuals in outcomes, and that outcomes \( x \) and \( y \) are such that \((w_1(x), \ldots, w_N(x)) = w \) and \((w_1(y), \ldots, w_N(y)) = v \). Then \( x \) at least as good as \( y \) iff \( w \succeq v \). Imagine, now, that we switch to \( \theta^*(\cdot) \) to assign well-being numbers to individuals in outcomes. The vector ranking remains the same. For example, \( \theta^*(w) \) such that \((w_1^*(x), \ldots, w_N^*(x)) = ww \) and \((w_1^*(y), \ldots, w_N^*(y)) = vv \). Then \( x \) at least as good as \( y \) iff \( ww \succeq vv \).

In the analytic set-up here, with an individual’s well-being determined by her history—\( w_i(x) = \theta(a_i(x), R_i(x)) \)—profile-independence means, more specifically, that \( \succeq \) is independent of \( \theta(\cdot) \). The ranking of well-being vectors remains the same, regardless of which \( \theta(\cdot) \) is used to assign well-being numbers to histories.

To say that a prioritarian SWF is profile-independent means that the choice of \( g(\cdot) \) is independent of \( w(\cdot) \)—specifically, independent of \( \theta(\cdot) \). Putting together the requirement of profile-independence with invariance considerations, we arrive at the Atkinson family of prioritarian SWFs. Atkinson SWFs define the \( g(\cdot) \) function as follows: \( g(w_i) = \frac{1}{1-\gamma} w_i^{1-\gamma} \), with \( \gamma > 0 \) (and \( \gamma \neq 1 \)), and \( g(w_i) = \log w_i \) in the special case of \( \gamma = 1 \).

---

76 \( w^*(\cdot) \) is a “common cardinal rescaling” of \( w(\cdot) \) if there exists \( k > 0, l, s.t. w_i^*(x) = kw_i(x) + l \) for all \( i, x \). \( \theta^*(\cdot) \) is a “positive affine transformation” of \( \theta(\cdot) \) if there exists \( r > 0, s \) s.t. \( \theta^*(h) = r\theta(h) + s \) for every \( h \). Consider any \( \theta^*(\cdot) \) which is a positive affine transformation of \( \theta(\cdot) \), and the corresponding \( w^*(\cdot) \): \( w_i^*(x) = \theta^*(h_i(x)) = \theta^*(h(x)) + s = rw_i(x) + s \). So \( w^*(\cdot) \) is indeed a common cardinal rescaling of \( w(\cdot) \), with \( k = r \) and \( l = s \).
No profile-independent prioritarian SWF is invariant to a common cardinal rescaling of \( w(\cdot) \). If we relax the invariance requirement somewhat—requiring invariance to a common ratio rescaling of \( w(\cdot) \)—this impossibility is avoided. If the SWF is profile-independent, prioritarian, and invariant to a common ratio rescaling of \( w(\cdot) \), then it must be an Atkinson SWF.

In the set-up here, requiring that the SWF be invariant to a common ratio rescaling of \( w(\cdot) \) is equivalent to requiring that \( \theta(\cdot) \) be unique up to a positive ratio transformation.\(^{77}\) That is to say, \( \theta(\cdot) \) should contain information not only about the well-being levels of histories and well-being differences, but also about well-being ratios. In effect, we need to take \( \theta(\cdot) \), which already represents levels and differences, and now “add in” information about well-being ratios.

We do so by picking a zero history \( h^\text{zero} = (a^\text{zero}, R^\text{zero}) \). Conceptually, well-being ratios are just ratios of differences from a zero point. The well-being ratio between history \( h \) and history \( h^* \) is just the ratio of the well-being difference between \( h \) and a zero point \( h^\text{zero} \), and the well-being difference between \( h^* \) and that zero point. With \( h^\text{zero} \) in hand, we can now specify a new well-being measure \( \theta^*(\cdot) \), with \( \theta^*(a, R) = \theta(a, R) - \theta(a^\text{zero}, R^\text{zero}) \). Note that \( \theta^*(\cdot) \) preserves all of the well-being level and difference information in \( \theta(\cdot) \), but also now represents well-being ratios; and is unique up to a positive ratio transformation.

How should \( h^\text{zero} \) be chosen? Relevant considerations are the following. (1) We might have intuitions about the well-being ratios between histories. Such intuitions can guide the choice of \( h^\text{zero} \). For example, assume that \( \theta(a, R) > \theta(a^+, R^+) \), and that we intuit that someone with history \( (a, R) \) is twice as well off as someone with history \( (a^+, R^+) \). Then we can choose \( h^\text{zero} = (a^\text{zero}, R^\text{zero}) \) so that \( [\theta(a, R) - \theta(a^\text{zero}, R^\text{zero})] / [\theta(a^+, R^+) - \theta(a^\text{zero}, R^\text{zero})] = 2. \) (2) A strong pragmatic consideration is that the well-being numbers inputted into the Atkinson SWF must be non-negative (or positive with \( \gamma \geq 1 \)).\(^{78}\) Thus, \( h^\text{zero} \) should be chosen so that every well-being vector that might arise with non-zero probability contains only non-negative (or positive, with \( \gamma \geq 1 \)) entries. That is: \( h^\text{zero} \) should be a sufficiently “bad” history that, for every outcome \( y \) such that \( x_P, s = y \) for some policy \( P \) in \( \mathbf{P} \) and some state \( s \), and for every \( i \), \( \theta(a_i(y), R_i(y)) \geq \theta(a^\text{zero}, R^\text{zero}) \) —or \( \theta(a_i(y), R_i) > \theta(a^\text{zero}, R^\text{zero}) \) for \( \gamma \geq 1 \). (3) \( h^\text{zero} \) is the point of absolute moral priority, in the following sense. As an individual’s well-being level approaches \( h^\text{zero} \), the relative weight given to changes in her well-being by an Atkinson SWF, as compared to that of better-off individuals, approaches infinity. For any prioritarian SWF, if individual \( i \) is at well-being level \( w_i \), the marginal moral impact of a change to her well-being is \( g'(w_i) \). For an Atkinson SWF, \( g'(w_i) = w_i^{-\gamma} \). Consider two individuals \( i \) and \( j \), \( 0 < w_i < w_j \). According to the Atkinson SWF, the marginal moral impact of a change to \( i \)’s well-being, as a multiple of the marginal moral

\(^{77}\) \( w^*(\cdot) \) is a “common ratio rescaling” of \( w(\cdot) \) if there exists \( k > 0 \) s.t. \( w_i^*(x) = kw_i(x) \) for all \( i, x \). \( \theta^*(\cdot) \) is a “positive ratio transformation” of \( \theta(\cdot) \) if there exists \( r > 0 \) s.t. \( \theta^*(h) = r\theta(h) \). Following an analysis parallel to that in note above, it’s clear that if \( \theta^*(h) = r\theta(h) \), then the corresponding \( w^*(\cdot) \) is such that \( w_i^*(x) = kw_i(x) \), with \( k = r \).

\(^{78}\) As explained in the main text, the Atkinson \( g(\cdot) \) function is either undefined or, if defined, not both strictly increasing and strictly concave with negative well-being numbers as inputs (or with 0 as an input for \( \gamma \geq 1 \)).
impact of a change to \(j\)'s, is \(\frac{g'(w_j)}{g'(w_i)} = \left(\frac{w_j}{w_i}\right)^\gamma\). For any value of \(\gamma > 0\), this ratio approaches infinity as \(w_i\) approaches 0.

Atkinson SWFs, again, are a family of SWFs. Picking a value of \(\gamma\) identifies a particular SWF. How should \(\gamma\) be chosen? \(\gamma\) is a priority parameter.\(^79\) As the equation immediately above shows, \(\gamma\) specifies the marginal moral impact of a change to a worse-off individual’s well-being, as a multiple of the marginal moral impact of a change to a better-off person’s well-being, depending on the ratio of their well-being levels (better-off to worse-off). If \(w_j > w_i > 0\) and \((w_j/w_i) = K\), then this ratio of marginal moral impacts is \((w_j/w_i)^\gamma = K^\gamma\).

With \(\gamma = 0\), the Atkinson SWF is no longer prioritarian but becomes utilitarian. (This can be seen immediately by plugging \(\gamma = 0\) into the formula above.) As \(\gamma\) increases, the Atkinson SWF gives increasing priority to the worse-off: the ratio of marginal moral impacts \((w_j/w_i)^\gamma = K^\gamma\) increases with \(\gamma\) for a given well-being ratio \(K\) between the better- and worse-off individuals.

The ratio of marginal moral impacts can also be thought of in terms of leaky transfers. Imagine reducing the better-off individual’s well-being by some small amount \(\Delta w\) and increasing the worse-off individual’s well-being by a fraction of the loss to the better-off one, i.e., by \(\rho\Delta w\). 0 < \(\rho\) < 1. The moral impact of the loss to the better-off one is approximately \(- (\Delta w) w_j^{-\gamma}\), while the moral impact of the gain to the worse-off one is approximately \((\rho\Delta w) w_i^{-\gamma}\). Let \(\rho_{min} = (1/K)\). Note that \((\rho\Delta w) w_i^{-\gamma} < (\Delta w) w_j^{-\gamma}\) iff \(\rho < \rho_{min}\). So we can test whether the leaky transfer is a net moral gain or loss by seeing whether fractional gain \(\rho\) to the worse-off one is greater than or less than \(\rho_{min}\). This is an approximate test that becomes increasingly accurate as \(\Delta w\) approaches 0.

\(\rho_{min}\) is (approximately) the minimum acceptable fraction for a leaky transfer. Intuitions about \(\rho_{min}\) for a given \(K\) can be used to deliberate about the appropriate level of \(\gamma\). Note that \(\gamma = -(\log \rho_{min}/\log K)\).

Having chosen a value of \(\gamma\), we now use the resultant Atkinson SWF together with a well-being measure \(w^*(\cdot)\) that is unique up to a common ratio rescaling—with \(w^*(\cdot)\) specified, as discussed above, by taking an \(\theta(\cdot)\) that represents well-being levels and differences; identifying an \(h_{zero} = (a_{zero}, R_{zero})\); defining \(\theta^*(a, R) = \theta(a, R) - \theta(a_{zero}, R_{zero})\); and setting \(w_i^*(x) = \theta^*(a_i(x), R_i(x))\). That is, we rank outcomes as follows—\(x\) at least as good as \(y\) iff
\[
\frac{1}{1/(1-\gamma)} \sum_{i=1}^{N} w_i^*(x)^{1-\gamma} \geq \frac{1}{1/(1-\gamma)} \sum_{i=1}^{N} w_i^*(y)^{1-\gamma}
\]
— and we rank policies by using our chosen uncertainty module with \(w^*(\cdot)\) as the well-being measure. For example, the EPP score assigned to a policy \(P\) would be:
\[
\sum_{s \in S} \pi(s) (1/(1-\gamma)) \sum_{i=1}^{N} w_i^*(x_{s,i})^{1-\gamma}.
\]

\(^{79}\) A standard mathematical indicator of the degree of concavity of a concave function \(g(m)\) at a value \(m\) is the coefficient of relative risk aversion: \(-g''(m)mg'(m)\). (Lambert 2001, pp. 94-97). In the case of the Atkinson \(g(\cdot)\) this coefficient is constant and equal to \(\gamma\).
This approach is profile-independent because $\gamma$ is independent of $w^*(\cdot)$, i.e., independent of $\theta^*(\cdot)$. Note that $\gamma$ is not a function of $\theta(\cdot)$ and $h^{\text{zero}}$. Note also that, in the leaky-transfer thought experiment suggested above for specifying $\gamma$, the choice of $\gamma$ is meant to depend upon intuitions about $\rho_{\text{min}}$ for a given well-being ratio $K = w_j/w_i$. It is not meant to depend upon which well-being measure $\theta^*(\cdot)$ is used to map the two individuals’ histories onto well-being numbers and thereby to determine the well-being ratio between them.

2.A.3.2 Kolm-Pollak SWF (Profile-Dependent)

Kolm-Pollak SWFs are a family of prioritarian SWFs that define the $g(\cdot)$ function as follows: $g(w_i) = -\exp(-\beta w_i), \beta > 0$. As compared to Atkinson SWFs, Kolm-Pollak SWFs have the advantage of allowing well-being numbers to take any value: positive, zero, or negative. Moreover, unlike Atkinson SWFs, which accord priority to a worse-off individual approaching infinity as her well-being level approaches 0, Kolm-Pollak SWFs do not have this sort of property (a property that is intuitively troubling). With individual $i$ at well-being level $w_i$, the marginal moral impact of a change to $i$’s well-being is $\beta \exp(-\beta w_i)$. With $w_i < w_j$, the ratio of marginal moral impacts between the worse-off and better-off individuals is $\exp(\beta(w_j - w_i))$, which is finite for any number $w_i$ (whether $w_i$ is positive, negative, or zero) and approaches infinity only as $(w_j - w_i)$ does.

With $\beta = 0$, the Kolm-Pollak SWF is no longer prioritarian but becomes utilitarian. As $\beta$ increases, the Kolm-Pollak SWF gives increasing priority to the worse-off: the ratio of marginal moral impacts $\exp(\beta(w_j - w_i))$ increases with $\beta$ for a given well-being difference $(w_j - w_i)$ between the better- and worse-off one.

Kolm-Pollak SWFs are not invariant to a common cardinal rescaling of $w(\cdot)$ (no prioritarian SWFs are), nor are they invariant to a common ratio rescaling of $w(\cdot)$ (only Atkinson SWFs are)—but Kolm-Pollak SWFs do have the property of being the only prioritarian SWFs that are invariant to a common translation rescaling of $w(\cdot)$. (Bossert and Weymark 2004, pp. 1160-61). To say that $w^*(\cdot)$ is a common translation rescaling of $w(\cdot)$ means: $w^*_i(x) = w_i(x) + l$ for all $i, x$. With $w(\cdot)$ derived from $\theta(\cdot)$, requiring that $w(\cdot)$ be unique up to a common translation rescaling is equivalent to requiring that $\theta(\cdot)$ be unique up to a translation. $\theta^*(\cdot)$ is a translation of $\theta(\cdot)$ if $\theta^*(h) = \theta(h) + s$ for every history $h$.

Note that $\theta(\cdot)$ and all positive affine transformations thereof do not agree about the magnitude of the well-being difference between histories. However, $\theta(\cdot)$ and all translations therefore do agree about that magnitude. If $\theta(h') - \theta(h) = D$, and $\theta^*(\cdot)$ is a translation of $\theta(\cdot)$, then $\theta^*(h') - \theta^*(h) = D$.

---

80 Note that the ratio of marginal moral impacts, $\exp(\beta(w_j - w_i))$, is 1 with $\beta = 0$. Note also that the Kolm-Pollak SWF is ordinarily equivalent to summing $g^*(w_i) = (-\exp(-\beta w_i) + 1)/\beta$. By l’Hôpital’s rule, the limit of $g^*(\cdot)$ as $\beta$ approaches 0 is $w_i$.
81 Consider $\theta^*(\cdot) = r\theta(\cdot) + s$, with $r \neq 1$. If $\theta(h') - \theta(h) = D$, $\theta^*(h') - \theta^*(h) = rD \neq D$. 

68
In principle, we could combine a profile-independent Kolm-Pollak SWF with a well-being measure \( \theta(\cdot) \) unique up to a translation, as follows. (1) Starting with \( \theta(\cdot) \), which represents our comparisons of well-being levels and differences, we “add in” information about the magnitude of well-being differences. Assume that we judge that the well-being difference between \( h' \) and \( h \) is \( D \), while in fact \( \theta(h') - \theta(h) = d \). Then let \( r = D/d \), and let \( \theta^+(\cdot) = r\theta(\cdot) \). Then \( \theta^+(\cdot) \) reflects this well-being information, as do all translations of \( \theta^+(\cdot) \). (2) The priority parameter for the Kolm-Pollak SWF is \( \beta \). We specify a profile-independent value for \( \beta \)—a value that is independent of how we choose \( \theta(\cdot) \) and \( r \)—by deliberating about how the relative marginal moral impact of changes to the well-being of two individuals, \( w_i \) and \( w_j \), with \( w_i < w_j \), depends upon the magnitude of the well-being difference between them. As already stated, this ratio of marginal moral impacts is just \( \exp(\beta(w_j - w_i)) \). The minimum acceptable fraction for a leaky transfer is approximately \( \rho_{\text{min}} = \exp(\beta(w_i - w_j)) \).

The difficulty with the approach sketched in the previous paragraph is that it’s unclear what the magnitude of a well-being difference means. We can coherently judge that Philippe is better off than Suzanne, or that the well-being difference between Philippe and Suzanne is greater than the well-being difference between Ariela and Juan. But it doesn’t seem meaningful to say that the well-being difference between Philippe and Suzanne is 7 rather than 5 or 110.

A more feasible approach, which doesn’t require judgments about the magnitude of well-being differences, is to specify a profile-dependent Kolm-Pollak SWF so as to fit judgments regarding the moral weight of changes to individuals’ attributes. Let \( a^m \) denote the attribute of type \( m \). Assume that individual \( i \) has history \((a^*, R^*)\), and individual \( j \) has history \((a^+, R^+)\), with \( i \) worse off than \( j \), i.e., \( \theta(a^*, R^*) < \theta(a^+, R^+) \). The well-being impact of a marginal change \( \Delta a^m \) to \( i \)’s holdings of \( a^m \) is approximately \( \frac{\partial \theta}{\partial a^m}(a^*, R^*)\Delta a^m \), while the well-being impact of a marginal change \( \Delta a^m \) to \( j \)’s holdings of \( a^m \) is approximately \( \frac{\partial \theta}{\partial a^m}(a^+, R^+)\Delta a^m \). Using the Kolm-Pollak SWF, the moral impact of the first change is \( \exp(-\beta(\theta(a^+, R^+) - \theta(a^*, R^+))) \frac{\partial \theta}{\partial a^m}(a^*, R^*)\Delta a^m \), and that of the second is \( \exp(-\beta(\theta(a^+, R^+) - \theta(a^*, R^+))) \frac{\partial \theta}{\partial a^m}(a^+, R^+)\Delta a^m \), so that the ratio of marginal moral impacts,

\[
\frac{\frac{\partial \theta}{\partial a^m}(a^+, R^+)}{\frac{\partial \theta}{\partial a^m}(a^*, R^*)} = Z.
\]

intuitions about the magnitude of this ratio \( (Z) \) to solve for \( \beta \).

\[
\beta = \left[ \log Z + \log \frac{\partial \theta}{\partial a^m}(a^+, R^+) - \log \frac{\partial \theta}{\partial a^m}(a^*, R^*) \right] / \left[ \theta(a^+, R^+) - \theta(a^*, R^+) \right].
\]

Note that this approach is profile-dependent because \( \beta \) depends upon \( \theta(\cdot) \).
The preceding is quite abstract, and a more concrete example may help. Some of the literature on SWFs contemplates using thought experiments about leaky transfers of income to specify an SWF. Assume that all histories have the same preference, so that \( \theta(a, R) \) becomes \( \theta(a) \). Let \( c \) denote income and \( b \) other attributes, so that \( \theta(a) = \theta(c, b) \). Assume that income has a positive well-being impact: \( \frac{\partial \theta}{\partial c} > 0 \). Assume that individual \( i \) has income \( c^* \), and individual \( j \) has income \( c^+ > c^* \), with both individuals’ non-income attributes at level \( b \). Then the ratio between the marginal moral impact of a change \( \Delta c \) to \( i \)’s income and a change \( \Delta c \) to \( j \)’s income is

\[
\frac{\exp(\beta(\theta(c^+, b) - \theta(c^*, b)))}{\frac{\partial \theta}{\partial c}(c^+, b)} = Z.
\]

Intuitions about this ratio are just intuitions about the minimum acceptable fraction \( f_{\text{min}} \) for a leaky transfer in income from individual \( j \) to individual \( i \).

\[ f_{\text{min}} = 1/Z. \]

**2.A.3.3 Atkinson SWF (Profile-Dependent)**

A third possibility is to have a profile-dependent Atkinson SWF. Since the Atkinson and Kolm-Pollak functional forms are the two most widely used types of \( g(\cdot) \) functions, and since we have just discussed how to specify a profile-dependent Kolm-Pollak SWF, this possibility also should be mentioned.

In specifying a profile-dependent Atkinson SWF, there is no need to construct a \( w^*(\cdot) \) unique up to a common ratio rescaling by starting with \( \theta(\cdot) \), identifying a zero history \( h^\text{zero} \), and thereby arriving at \( \theta^*(\cdot) \) unique up to a positive ratio transformation and at \( w^*(\cdot) \) defined from \( \theta^*(\cdot) \).

Rather, we simply pick \( \gamma \) given \( \theta(\cdot) \) so as to fit intuitions about the relative moral impact of changes to individuals’ attributes.

As with the Kolm-Pollak analysis above, let \( a^m \) denote the attribute of type \( m \). Assume that individual \( i \) has history \( (a^*, R^*) \), and individual \( j \) has history \( (a^+, R^+) \). With \( \theta(\cdot) \) in hand, the well-being impact of a marginal change \( \Delta a^m \) to \( i \)’s holdings of \( a^m \) is approximately

\[
\frac{\partial \theta}{\partial a^m}(a^*, R^*)\Delta a^m,
\]

while the well-being impact of a marginal change \( \Delta a^m \) to \( j \)’s holdings of \( a^m \) is approximately

\[
\frac{\partial \theta}{\partial a^m}(a^+, R^+)\Delta a^m.
\]

Now using the Atkinson SWF, the moral impact of the first change is

\[ Z. \]

---

82 This is only needed for the profile-independent Atkinson SWF. That SWF is not invariant to a common cardinal rescaling, but is invariant to a common ratio rescaling.

83 The profile-dependent Atkinson SWF does still require well-being numbers to be non-negative (or positive with \( \gamma \geq 1 \)). If \( \theta(\cdot) \) is such that well-being vectors with negative entries can arise, define \( \theta^*(\cdot) \) as \( \theta(\cdot) + t \), with \( t \) a sufficiently large positive number that all well-being numbers are positive: For every \( i \), and for every \( y \) such that \( x_{P,s} = y \) for some state \( s \) and some policy \( P \) in \( P \), \( \theta^*(a(y), R(y)) > 0 \). Then undertake the exercise that follows in terms of \( \theta^*(\cdot) \), i.e., choose \( \gamma \) given \( \theta^*(\cdot) \) to fit intuitions about the relative moral impact of attribute changes.
change is \( \theta(a^*, R^*)^{-\gamma} \frac{\partial \theta}{\partial a_m^m}(a^*, R^*) \Delta a_m^m \), and that of the second is \( \theta(a^+, R^+)^{-\gamma} \frac{\partial \theta}{\partial a_m^m}(a^+, R^+) \Delta a_m^m \),

so that the ratio of marginal moral impacts is:

\[
\left( \frac{\theta(a^+, R^+)}{\theta(a^*, R^*)} \right)^{\gamma} \frac{\partial \theta}{\partial a_m^m}(a^*, R^*) = Z. \]

We can use our intuitions about the magnitude of this ratio (Z) to solve for \( \gamma \).

\[
\gamma = \left[ \log Z + \log \frac{\partial \theta}{\partial a_m^m}(a^+, R^+) - \log \frac{\partial \theta}{\partial a_m^m}(a^*, R^*) \right] / \left[ \log \theta(a^+, R^+) - \log \theta(a^*, R^*) \right].
\]

References


