Environ Resource Econ (2015) 62:279–308
DOI 10.1007/s10640-015-9960-7

Prioritarianism and Climate Change

Matthew D. Adler[1] · Nicolas Treich[2]

Accepted: 14 August 2015 / Published online: 11 September 2015

Abstract Prioritarianism is the ethical view that gives greater weight to well-being changes affecting individuals at lower well-being levels. This view is influential both in moral philosophy, and in theoretical work on social choice—where it is captured by a social welfare function ("SWF") summing a concave transformation of individual well-being numbers. However, prioritarianism has largely been ignored by scholarship on climate change. This Article compares utilitarianism and prioritarianism as frameworks for evaluating climate policy. It reviews the distinctive normative choices that are required for the prioritarian approach: specifying a ratio scale for well-being (if the prioritarian SWF takes the standard "Atkinson" form); determining the degree of concavity of the transformation function (i.e., the degree of social inequality aversion); and choosing between "ex ante" and "ex post" prioritarianism under conditions of risk. The Article also sketches some of salient implications of a prioritarian SWF for climate policy—with respect to the social cost of carbon, the social discount rate, optimal mitigation, and the "dismal theorem." Finally, it discusses the issue of variable population.

Keywords Climate change · Social choice · Social welfare function · Prioritarianism · Risk and uncertainty · Population

1 Introduction

Most research in climate economics can be seen as an exercise in ethics. A common way to deal with ethics in economics is through the specification of the "social welfare

---

July 2015.

✉ Matthew D. Adler
adler@law.duke.edu

1 Duke University School of Law, Durham, NC, USA

2 Toulouse School of Economics (LERNA, INRA), Toulouse, France

function" (SWF). The debates following the Stern Review (2007) about time discounting (Dasgupta 2008; Sterner and Persson 2008; Roemer 2010) and about the treatment of risk (Weitzman 2011; Millner 2013; Botzen and van den Bergh 2014) have illustrated the importance of the SWF specification. Recently, several papers have examined empirically the sensitivity of climate policy to the SWF (Llavador et al. 2011; Dietz and Asheim 2012).

Chapter 3 of Working Group III's Fifth Assessment Report of the IPCC is dedicated to ethics (IPCC 2014). This chapter, entitled *Social, Economic and Ethical Concepts and Methods*, devotes a section to the use of SWFs in climate economics. In this section, two kinds of SWFs are discussed: utilitarian and prioritarian SWFs. These SWFs have been commonly used in welfare and public economics (Boadway and Bruce 1984; Bossert and Weymark 2004; Drèze and Stern 1987; Kaplow 2008). However, most of the climate economics literature has considered only a utilitarian SWF.

Within academic moral philosophy, the concept of prioritarianism derives from Parfit (1991). There is now a substantial body of work by philosophers defending the prioritarian approach to moral thinking, and criticizing utilitarianism (Arneson 2007; Brown 2005; Holtug 2010, 2015; McKerlie 2007; Parfit 2012; Porter 2012; Tungodden 2003; Williams 2012). It is therefore important, we believe, for scholars to begin systematically exploring the implications of prioritarianism for policy choice with respect to climate change.

However (as shall emerge over the course of this Article) the *specification* of the prioritarian SWF as a methodology for evaluating climate policy involves a wide range of normative choices. Some of these normative choices (for example, the adoption of an interpersonally comparable well-being function) are *also* choices that the proponent of the utilitarian SWF must confront. Other choices (for example, the degree of inequality aversion with respect to well-being) are unique to prioritarianism.

In Part 2 of the Article, we describe the SWF concept and summarize the normative case—much more fully presented elsewhere—in favor of a prioritarian rather than utilitarian SWF. In the remainder of the Article, we review the range of normative questions that must be addressed in applying the prioritarian framework to the problem of climate change. These questions fall into four groups: the *measurement of well-being* (Part 3); the *degree of social inequality aversion* (Part 4); *social choice under risk* (Part 5); and *endogenous (variable) population* (Part 6). Throughout, we indicate the similarities and differences between the normative choices that arise in specifying the prioritarian SWF, and the choices that arise in specifying the utilitarian SWF.

In the course of our discussion, we also illustrate some of the *implications* of the prioritarian approach for climate policy. A comprehensive review of such implications is beyond the scope of this Article (since, of course, the implications depend upon the answers to the four groups of normative questions that we will survey). Still, we will point in a preliminary way to some of the clear implications of the prioritarian framework.

A central theme of the Article is that normative questions are inescapable when it comes to climate policy. Anyone who takes a stand in favor of a concrete policy, or in favor of some tool for evaluating policy, is making a normative commitment—implicit if not explicit. Science alone cannot justify such a commitment. It would be absurd to deny that good climate policy should take account of empirical facts; but it must also be stressed that the way in which appropriate climate policy depends upon the empirical facts is a function of the normative framework that the policymaker adopts.

2 The SWF Approach in Climate Economics

In this part of the Article, we first present the SWF concept. We provide classical references on that concept, and review some important preliminary points (regarding the determinants of well-being, and the role of a well-being discount factor) that hold true regardless of the functional form of the SWF—be it utilitarian, prioritarian, or other. We then summarize the normative case in favor of a prioritarian rather than utilitarian SWF.

2.1 The SWF Construct and Climate Economics

The SWF is a methodology for evaluating governmental policies that is widely used in theoretical welfare economics, optimal tax theory, optimal growth theory, environmental economics, and other areas of economics. For citations to these literatures and a general defense of the SWF methodology, see Adler (2012). The methodology starts with a set of possible *outcomes*. An "outcome" is a joint consequence. It describes what might happen as a result of the policy to everyone in the population of interest. We will use the symbols "x" and "y" to denote outcomes.

A given person's *attributes* will vary from outcome to outcome. By "attributes," we mean both someone's consumption (expenditure on marketed goods) and non-consumption attributes. We use the symbol $c_i(x)$ to denote individual i's consumption in outcome x; and the symbol $\mathbf{a}_i(x)$ to denote her non-consumption attributes (possibly a vector) in outcome x.

A given person's *preferences* can also vary from outcome to outcome, and certainly different individuals can have different preferences. Someone's "preference" is a ranking of consequences by that individual, a ranking which motivates and explains her choices. Formally, we think of a preference as a ranking of possible attributes—of $(c, \mathbf{a})$ combinations—and of lotteries over possible attributes. We use the symbol R to denote a preference, and $R_i(x)$ the preference of individual i in outcome x. For more discussion concerning preference and well-being, see Part 3.

We assume a finite (rather than infinite) population of interest—an assumption that will be defended below. Moreover, we assume for now that the finite population has the same, fixed size N in all outcomes. This assumption is relaxed in Part 6, which addresses variable populations.

The SWF methodology employs an interpersonally comparable well-being function $v(.)$. $v(.)$ maps a given outcome onto a vector of well-being numbers, one for each person in the population. We will use $v(.)$ to denote a vector-valued function taking outcomes as arguments, and $v_i(x)$ the ith component of $v(x)$. We use $u(.)$ to denote a scalar-valued function taking bundles of individual attributes and preferences as arguments. $u(c, \mathbf{a}, R)$ measures the well-being associated with bundle $(c, \mathbf{a}, R)$.[1] $v(.)$ and $u(.)$ are logically related as follows: $v_i(x) = u(c_i(x), \mathbf{a}_i(x), R_i(x))$.

Thus, with a finite population of size N, outcome x is mapped onto the vector of well-being numbers $(v_1(x), \ldots, v_N(x)) = (u(c_1(x), \mathbf{a}_1(x), R_1(x)), \ldots, u(c_N(x), \mathbf{a}_N(x), R_N(x)))$.

Within the SWF tradition, $v(.)$ and $u(.)$ are commonly called "utility" functions. However, so as to emphasize that these functions are not necessarily the same as the ordinary utility functions that economists use—that further normative choices may be needed to construct $u(.)$ and $v(.)$—we will refer to them as *well-being* functions or measures.

[1] Indeed, well-being also depends upon the price vector; but we leave that implicit.

With these tools in hand, the SWF methodology ranks outcomes via some rule M for ranking vectors of well-being numbers. That is, the methodology says: outcome x is at least as good as outcome y iff $v(x)$ is ranked by M at least as good as $v(y)$. This outcome ranking, in turn, generates a ranking of the *policies* (choices) available to the decisionmaker. In the simplest case, where the decisionmaker knows with certainty which outcome a given choice will produce, the rule for doing so is trivial.[2] Using the SWF framework to rank choices with risky outcomes raises more complicated questions, addressed in Part 5.

A long tradition, going back to Bergson (1948, 1954), Samuelson (1947 p. 221), and Harsanyi (1977, Ch. 4) sees the SWF as a tool for *ethical/moral* deliberation. That is the perspective adopted here. More specifically, the SWF is a framework for *welfarist* ethical deliberation. Welfarists see individual well-being as the foundation for ethical thought: if two outcomes are identical with respect to everyone's well-being, they are equally ethically good. Non-welfarist ethical approaches—for example, "deontological" views—are not well captured by the SWF framework.

Every plausible SWF, understood as a tool for welfarist ethical deliberation, satisfies two basic axioms: the (strong) Pareto principle, and anonymity. The Pareto principle is the touchstone of welfare economics. Expressed as a constraint on SWFs, it says that if $v_i(x) \geq v_i(y)$ for all i, with the inequality strict for some, then $v(x)$ is strictly preferred by M to $v(y)$. The axiom of anonymity captures the attitude of impartiality that is fundamental to ethics. This axiom says: if $v(x)$ is a permutation of $v(y)$, then M is indifferent between the two. The idea here is that if two outcomes give rise to one and the same pattern of well-being levels, differing only in the identities of the individuals who end up at the various levels, then the outcomes are ethically equally good. On axiomatic characterizations of SWFs, see Adler (2012, Ch. 5), Boadway and Bruce (1984, Ch. 5), Blackorby et al. (2005), Bossert and Weymark (2004), Weymark (2016).

A very wide array of SWFs satisfy these two basic axioms. One is the utilitarian SWF, which takes the form: outcome x at least as good as outcome y iff $\sum_{i=1}^{N} v_i(x) \geq \sum_{i=1}^{N} v_i(y)$. The prioritarian SWF sums well-being numbers "transformed" by a strictly concave, strictly increasing function $g(.)$. It says: outcome x at least as good as outcome y iff $\sum_{i=1}^{N} g(v_i(x)) \geq \sum_{i=1}^{N} g(v_i(y))$.[3] Note that both the utilitarian and prioritarian SWFs are social welfare *functions* in the strict sense that they use a real-valued function to encapsulate the rule M for ordering vectors of well-being numbers. Abbreviate this function as $w(.)$. For prioritarians, we have[4] $w(x) = w(v(x)) = \sum_{i=1}^{N} g(v_i(x))$.

Although some economists are skeptical about SWFs (preferring instead a criterion of Kaldor–Hicks efficiency), this is *not* true of economists working on climate change. The SWF methodology is pervasive in this literature. For example, both the *Stern Review* (Stern 2007) and Nordhaus' book *A Question of Balance* (2008) employ an SWF as the fundamental tool for evaluating carbon-reduction policies. Their policy differences are profound—but because of a disagreement about the *parameters* of the SWF (in particular, the discount rate), and not because of a disagreement about the methodology itself.

[2] Let $x(a)$ denote the certain outcome of choice a. Then a at least as good as b iff $v(x(a))$ is ranked by M at least as good as $v(x(b))$.

[3] Strictly speaking, the prioritarian SWF use a well-being function that has been "zeroed out" or otherwise rescaled. Following the discussion of this issue in Part 3.2, we will use the symbols $u^*(.)$ and $v^*(.)$ to denote such a well-being function, but do not do so here since the difference between these and the $u(.)/v(.)$ functions has not yet been explained.

[4] We use "$w(x)$" as a shorthand for $w(v(x))$.

We make some preliminary observations about existing climate scholarship. First, an individual's consumption (or income) is often used as the sole argument for the well-being function. That is, $v_i(x) = u(c_i(x))$, with $c_i(x)$ the consumption of individual i in this outcome. Second, the functional form of the SWF is often *discounted utilitarianism*. One standard version of this formula assumes discrete time and a final period, so that $w(x) = \sum_{t=0}^{T\max} (1+\rho)^{-t} N_t u(c_t(x))$, with ρ a well-being discount rate; N_t the number of individuals in the generation at time t; and $c_t(x)$ the consumption of everyone in that generation in x, assumed to be identical. With continuous time, this becomes $\int_{t=0}^{T\max} e^{-\rho t} N_t u(c_t(x))dt$. A different version of discounted utilitarianism often employed by climate scholars allows for infinite time: either $\sum_{t=0}^{\infty} (1+\rho)^{-t} N_t u(c_t(x))$ or $\int_{t=0}^{\infty} e^{-\rho t} N_t u(c_t(x))dt$. Variations of these discounted-utilitarian forms that allow for intragenerational heterogeneity are also used. See Botzen and van den Bergh (2014), describing the prevalence of discounted utilitarianism in economic analysis of climate change.

Let us leave aside for the moment the use of a utilitarian SWF. There are two, deeper, difficulties with the formulas above.

First, an individual's consumption—her expenditure on marketed goods—is only *one* determinant of her well-being. Non-consumption goods such as health, leisure, happiness, social relations, environmental quality, etc. are also important sources of individual welfare. Moreover, as already noted, individuals have heterogeneous preferences with respect to tradeoffs between such goods, or between them and consumption. An expanded $u(.)$ of the form $u(c, \mathbf{a}, R)$ captures the combined role of all three sources of individual welfare. The simple $u(c)$ form employed in the formulas above should be seen as a modeling choice which enhances tractability at the expense of realism. It assumes, in effect, that individuals are homogeneous with respect to non-consumption goods and preferences, which is certainly not in fact true.[5]

Second, the SWF formulas above include a well-being discount rate, represented as ρ. This is commonly referred to in the climate-change literature as "utility" discounting, or "pure time preference."[6] We concur with the opponents of well-being/utility discounting, beginning with Ramsey (1928), who argue that it is "ethically indefensible." See Dasgupta (2012), describing scholars who have taken this view, and Arrow et al. (2014). It embodies a clear violation of the attitude of *impartiality* that is foundational to ethics. Well-being discounting gives priority to present over future individuals just because of the arbitrary fact that present individuals come into being earlier in time, or closer to the present. Indeed, the axiomatic expression of impartiality—the anonymity axiom—and well-being discounting are flatly inconsistent.[7]

Four counterarguments have been advanced to support well-being discounting. (1) *Sacrificing the present for the future*. It has been argued that utilitarianism without a pure rate of time preference, together with a positive rate of return on investment, may have an *ethically* unpalatable result: namely, that it can maximize the sum of (undiscounted) well-being across

5 On the importance of non-consumption attributes see Sterner and Persson (2008), Roemer (2010), and Neumayer (2013). On the possibility of normalizing consumption to take account of non-consumption attributes, see footnote 17.

6 Not to be conflated with consumption discounting.

7 Assume that the numbering of individuals corresponds to birth order, so that individual 1 comes into existence before individual 2, and so on. Consider well-being vector $v = (v_1, \ldots, v_N)$, such that $v_1 < v_2 < \cdots < v_N$; and let v^+ be a permutation of v such that $v_1^+ > v_2^+ > \cdots > v_N^+$. Then anonymity requires that the SWF be indifferent between the two vectors, but time-discounted utilitarianism prefers v^+.

generations to reduce present consumption to a very low level, instead investing present resources for future benefit (Farber 2003; Nordhaus 2007, 2008; Weitzman 2007). Such a result is indeed ethically unpalatable, but is partly mitigated within utilitarianism via a well-being function that is strictly concave in consumption. Prioritarianism has yet more powerful tools to avoid sacrificing the present for the future. The more concave the $g(.)$ function, the closer the prioritarian optimum approaches to welfare equality between the generations even with a positive rate of return on investment. See Part 4.

(2) *Infinite Time.* Without a pure rate of time preference, infinite sums or integrals of well-being may be undefined. Indeed, an axiomatic tradition in economics beginning with Koopmans (1960) shows that seemingly plausible axioms for ordering infinite well-being streams may require positive time preference.

However, these axioms may not on reflection be very compelling, and relaxing them may allow for a form of non-discounted utilitarianism or prioritarianism even with infinite time—for example, by using Weizsacker's (1965) "overtaking" criterion to rank infinite well-being streams (Adler 2009). More fundamentally, infinite time is a modeling apparatus that departs from reality. It seems exceedingly unlikely that the human species will continue to exist ad infinitum. To be sure, economists often sacrifice realism in their models for the sake of mathematical tractability. But we should be very suspicious of using an unrealistic model (here, infinite time) as the basis for our ethical deliberations, if the model "forces" us to adopt ethically indefensible views (pure time preference) when a less tractable but more realistic model (finite time) does not have these troubling implications. Moreover, since it is very hard to see how the size of the human population could become infinite in finite time, we believe an assumption of a finite population is also most defensible.

(3) *Extinction Risk.* Utility discounting has been defended by Stern (2007) as capturing the non-zero risk that humanity will cease to exist within any given time period. However, this approach to extinction risk treats it as exogenous. Since catastrophic climate change itself threatens the very existence of the human species, it would be better to allow extinction risk to be endogenous.

Moreover, the use by Stern (2007) of a discount factor to reflect extinction risk obscures the difficult normative issues posed by such risk. Extinction risk is simply one reason (among others) that population size is variable, not fixed. Possible outcomes are heterogeneous with respect to the total number of past, present, and future individuals—because the date at which humanity ceases to exist is not the same in all outcomes, *and* for many other reasons. Moreover, the outcomes of a given *policy* may be risky or uncertain (in part because extinction at a given date is risky or uncertain, but for many other reasons too!). As we discuss below, in Part 5 and Part 6, there are different plausible non-discounted SWFs for ranking outcomes with a variable population (such as total, average, or critical-level utilitarianism or prioritarianism) and for handling risk and uncertainty. An explicit treatment of each policy as risky or uncertain with respect to outcomes varying in the total size of population—with these policies then ranked by some non-discounted SWF together with some rule for social choice under risk or uncertainty—not only permits extinction risk to be endogenous, but is much more transparent with respect to the underlying normative choices.

(4) *Observed Policies.* Observed governmental policies do not seem to be consistent with undiscounted utilitarianism. If the present generation were maximizing the intergenerational sum of well-being, without any time preference, we would observe a much greater rate of social investment (Nordhaus 2007, 2008; Weitzman 2007).

This argument is implicit in the use of the Ramsey formula to "infer" a non-zero time preference from the observed rate of return on capital (Nordhaus 2007, 2008), the so-called "descriptive" approach to time discounting (Arrow et al. 2014). But (as David Hume famously observed) we cannot infer “ought” from “is.” The observation that the present generation “is” partial to its own interests does not imply that it “ought” to be partial as an ethical matter.

Here, it should be stressed that the ethical norm of impartiality is a *demanding* one. Individuals, in their actual behavior, almost always fall short of that norm. Someone who lived her day-to-day life in conformity with an SWF satisfying the anonymity axiom would be a kind of saint. Similarly, it would be extraordinary for a government to be fully impartial between its current citizens and others (the unborn or citizens of other countries).

There is no need to see these pervasive departures from fully ethical behavior as irrational. The great philosopher Sidgwick (1907) took the position that a decisionmaker might rationally choose to advance moral aims, *or* her own interests, or some mix of these considerations. Sidgwick used the term “prudence” to describe what maximizes the decisionmaker’s own well-being. On the extreme version of the Sidgwickian view, it would be *rational* for the present generation *either* to be fully ethical (maximizing an SWF without discounting), *or* to be fully prudent (to ignore entirely the interests of future generations), *or* to do anything “in between.” Even if the extreme Sidgwickian position is rejected, it is surely true that an individual or government may rationally depart from the ethical norm of impartiality to some substantial extent (Schelling 1995).[8]

The SWF construct (as we understand it here) offers a framework for fully ethical choice. That is why we (along with many in the SWF tradition) embrace the anonymity axiom. The SWF framework does not purport to guide an individual or government in making rational tradeoffs *between* ethics and prudence—tradeoffs that anyone except a saint will find herself making (Scheffler 1982; Kagan 1989). This is one reason, among others, why the structure of an SWF cannot be observed from real-world governmental or individual behavior.

In this Article, we focus on what an impartial, ethical perspective requires of climate policy, as formalized via an anonymous SWF. But we recognize that actual decisionmakers will almost always be motivated by some mixture of ethical and non-ethical considerations, and indeed we believe (along with Sidgwick) that mixed motivations of this sort are quite rational. The recommendations that follow from the SWF construct are, in our view, *one* input into the climate decisionmaker’s rational calculus.[9]

2.2 Utilitarian Versus Prioritarian SWFs

Given the preceding discussion, our focus throughout the Article will be on an SWF with no well-being discount factor and with finite time and population. For simplicity, we will present formulas in discrete rather than continuous time. Leaving aside until Part 6 the complexities of a variable population, we return to the formulas for a utilitarian SWF, $w(x) = \sum_{i=1}^{N} v_i(x)$, and for a prioritarian SWF, $w(x) = \sum_{i=1}^{N} g(v_i(x))$. If we assume (as climate scholars often do) that the world is divided into D groups (typically, regions) and that individuals are homogeneous in their welfare-relevant attributes within a given group at a given time, the formulas become $w(x) = \sum_{t=1}^{T\max} \sum_{d=1}^{D} N_{d,t} u(c_{d,t}^x, \mathbf{a}_{d,t}^x, R_{d,t}^x)$ and $\sum_{t=1}^{T\max} \sum_{d=1}^{D} N_{d,t} g(u(c_{d,t}^x, \mathbf{a}_{d,t}^x, R_{d,t}^x))$, with $N_{d,t}$ the number of individuals in region d at time t and $c_{d,t}^x$, $\mathbf{a}_{d,t}^x$, and $R_{d,t}^x$ the consumption, non-consumption attributes, and preferences,

[8] Note that deontologists construe ethical impartiality in a very different manner. For welfare consequentialists, however, impartiality just is anonymity. It is well recognized that consequentialist ethics is much more demanding than deontological ethics.

[9] How climate decisionmakers should balance ethical and non-ethical (prudential) considerations is a vital topic, one barely addressed in the literature (Blackorby et al. 2000), but not one we can pursue here.

respectively, of those individuals. The simplest case of intragenerational homogeneity has $D = 1$.

Prioritarianism rarely appears in climate scholarship. The dominant IAMs all embed a utilitarian SWF.[10] Utilitarianism of some sort also underlies most theoretical work. Partha Dasgupta, in an influential article on climate change, writes: "Some ethicists have proposed an ethical theory they call 'prioritarianism,' which says that an increase in the well-being of a rich person … should be assigned less social value than the same increase in the well-being of a poor person…. I have not understood why such an *ad hoc* ethical principle should be awarded a name" (Dasgupta 2008, 146 n. 4; see also Broome 2008, rejecting prioritarianism).

But prioritarianism is hardly "ad hoc." One of us has elsewhere presented a lengthy defense of the prioritarian SWF (Adler 2012, ch. 5). The core of this case is the Pigou–Dalton principle. The Pigou–Dalton principle states an ethical preference for a non-leaky, rank-preserving transfer of well-being from a better-off to a worse-off individual, if no one else is affected. That is, it requires that outcome y be ethically preferred to outcome x if: $v_i(x) > v_j(x)$; $v_i(y) = v_i(x) - \Delta \geq v_j(y) = v_j(x) + \Delta, \Delta > 0$; and for all $k \neq i, j, v_k(x) = v_k(y)$. Under these conditions, j has a claim to y over x; i has a claim to x over y. Since j is strictly worse off in at least one of the outcomes and weakly worse off in both, and the well-being differences are identical, j's claim is stronger, and thus outcome y is ethically better on balance than x.

The utilitarian SWF, of course, violates the Pigou–Dalton principle. A non-leaky transfer of well-being leaves the sum total of well-being unchanged. The Pigou–Dalton principle is a powerful axiomatic tool, which makes precise how utilitarianism is insensitive to the distribution of well-being, and clarifies why such insensitivity is ethically troubling.

The prioritarian SWF is characterized by the combination of the two most fundamental axioms (Pareto and anonymity), plus Pigou–Dalton, plus two additional ethically plausible axioms: separability and continuity. To be sure, the last two axioms can be challenged; but such a challenge does not enter into the debate *between* utilitarianism and prioritarianism, since the utilitarian SWF also satisfies separability and continuity. See also Grant et al. (2010) for a different characterization of the prioritarian SWF.

The prioritarian SWF is, more precisely, a family of SWFs. Any $w(x)$ of the form stated above, with $g(.)$ some strictly increasing and strictly concave function, falls within this family. A popular subfamily consists in Atkinson/isoelastic SWFs, which take the form: $w(x) = (1-\gamma)^{-1} \sum_{i=1}^{N} v_i(x)^{1-\gamma}$. The inequality-aversion parameter γ can be any positive value. With $\gamma = 0$, the Atkinson SWF collapses to utilitarianism; as γ approaches infinity, the Atkinson SWF approaches leximin. In the special case of $\gamma = 1$, the Atkinson formula is $w(x) = \sum_{i=1}^{N} \log v_i(x)$.

Atkinson SWFs are the only SWFs that satisfy the five axioms that characterize prioritarianism, *plus* the axiom of ratio-rescaling-invariance.[11] An example of a non-Atkinson prioritarian SWF is the negative exponential SWF, $w(x) = \sum_{i=1}^{N} -\exp(-kv_i(x)),\ k > 0$.

We now survey the variety of normative issues that must be addressed in using the prioritarian SWF to evaluate climate policies. Some of these issues are common to both the utilitarian and prioritarian SWF. Issues that are unique to prioritarianism are indicated with a dagger (†).

[10] This is true of Nordhaus' DICE model (the framework for *A Question of Balance*) and the related RICE model, which unlike DICE takes account of intratemporal, interregional inequality (Nordhaus 2010). It is also true of the PAGE model underlying the *Stern Review*, and of the influential FUND model developed by Richard Tol (see Anthoff et al. 2009). See generally Botzen and van den Bergh (2014).

[11] This says that if $v^+(.) = av(.)$, a positive, then $v(x)$ is at least as good as $v(y)$ iff $v^+(x)$ is at least as good as $v^+(y)$. For more discussion, see Part 3.2

3 The Measurement of Well-Being

Utilitarianism and prioritarianism face a common normative question: how to construct an interpersonally comparable well-being function $v(.)$, notwithstanding the heterogeneity of preferences. Prioritarians, however, require a more precise scaling of the well-being function than utilitarians, and additional normative choices need to be made to achieve such scaling. We first describe the common question, then the issues specific to prioritarianism.

Throughout, we assume a preference-based view of well-being, which stipulates that: if a given person prefers having the attribute bundle $(c, \mathbf{a})$ to $(c^+, \mathbf{a}^+)$, then that person is better off with the first bundle. This assumption may itself be contested—for example, by those who hold a happiness view or an "objective good" view of welfare (Sumner 1996). We leave aside the possibility of contestation at this deep level, and stick to the preference view that is pervasive in economics.

3.1 Interpersonal Comparability

It is well-known that all SWFs require some degree of interpersonal comparability. Let $v^+(.)$ be related to $v(.)$ by individual-specific affine[12] transformations. That is, $v_i^+(x) = a_i v_i(x) + b_i$ for all x, with a_i positive. Note that $v^+(.)$ makes the very same *intrapersonal* comparisons of well-being levels and differences as $v(.)$, but implies different *interpersonal* comparisons.[13] It is easy to see that neither the utilitarian SWF nor the prioritarian SWF will be invariant to the substitution of $v^+(.)$ for $v(.)$. Each may rank outcomes differently using $v^+(.)$ rather than $v(.)$. Indeed, a social choice result related to Arrow's theorem shows that no plausible (Paretian and anonymous) SWF will be invariant to the substitution of $v^+(.)$ for $v(.)$ (Bossert and Weymark 2004; Weymark 2016).

Much SWF scholarship (not merely work on climate change) assumes that individuals have homogeneous preferences. Everyone has the same ordering of $(c, \mathbf{a})$ bundles and of lotteries over such bundles. That is, for all outcomes x and y, and all individuals i and j, $R_i(x) = R_j(y) = R$. Let $\varphi^R(.)$ be a von-Neumann/Morgenstern (vNM) utility function representing the common preferences R. Then it is straightforward to use $\varphi^R(.)$ to define the well-being function $u(.)$ and thus $v(.)$. That is, $u(c, \mathbf{a}) = \varphi^R(c, \mathbf{a})$, and $v_i(x) = u(c_i(x), \boldsymbol{a}_i(x))$.

By vNM theory, $\varphi^R(.)$ is unique up to an affine transformation (Kreps 1988). Thus $v(.)$ as defined in the previous paragraph is unique up to a *common* affine transformation, and we have achieved *inter*personal comparability of well-being levels and differences.[14]

Indeed, we see this approach in existing climate scholarship. Suppressing non-consumption attributes (as is usually done here), it is often assumed by climate scholars that $\varphi^R(.)$ is a vNM function of consumption of the constant relative risk aversion (CRRA) form, i.e.,

[12] Throughout, we use "affine transformation" to mean "positive affine transformation," i.e., multiplication by a *positive* number and the addition of some number. Similarly, "ratio transformation" is shorthand for "positive ratio transformation," i.e., multiplication by a positive number.

[13] For all *x, y, z, w*: If $i = j = k = l$ then (1) $v_i(x) \geq v_j(y)$ iff $v_i^+(x) \geq v_j^+(y)$, and (2) $v_i(x) - v_j(y) \geq v_k(z) - v_l(w)$ iff $v_i^+(x) - v_j^+(y) \geq v_k^+(z) - v_l^+(w)$. However, if it is not the case that $i = j = k = l$, then (1) and (2) do not necessarily hold true.

[14] Note that if $\varphi^{+R}(.)$ is also a vNM function representing R, then there exist a, b such that $\varphi^{+R}(.) = a\,\varphi^R(.) + b$ for all bundles, a positive. Let $v^+(.)$ be defined from $\varphi^{+R}(.)$. Then $v_i^+(.) = av_i(.) + b$, with a positive. Crucially, observe that a and b are not indexed by i; this is what makes $v^+(.)$ a *common* affine transformation of $v(.)$. A common affine transformation has the property that (1) $v_i(x) \geq v_j(y)$ iff $v_i^+(x) \geq v_j^+(y)$, and (2) $v_i(x) - v_j(y) \geq v_k(z) - v_l(w)$ iff $v_i^+(x) - v_j^+(y) \geq v_k^+(z) - v_l^+(w)$, whether or not $i = j = k = l$.

$\varphi^R(c) = (1-\alpha^R)^{-1}c^{1-\alpha^R}$ (Botzen and van den Bergh 2014; Pindyck 2013; on CRRA utility, see generally Gollier 2001, ch.2). Then α^R is the elasticity of marginal utility of consumption, or the coefficient of relative risk aversion, parameterizing $\varphi^R(.)$ and assumed to be common to all individuals. We can "estimate" α^R by looking to empirical evidence of risk preferences.

However, the suppression of non-consumption attributes, and on top of that the assumption of common preferences, are simplifications which should—where feasible—be replaced by a more complete model of well-being. Sterner and Persson (2008) amend the DICE model by using a utility function that depends upon both consumption and environmental quality, and find that the optimal emission path is highly sensitive to the parameter in the utility function capturing the elasticity of substitution between the two goods.

A well-being function that allows for heterogeneous preferences can no longer take the form $u(c, \mathbf{a})$. This is clear. Let R_i be the preferences of individual i in all outcomes, and R_j the preferences of individual j. Moreover, assume that R_i ranks bundle $(c, \mathbf{a})$ over bundle $(c^+, \mathbf{a}^+)$, while R_j has the opposite ranking. Given a preference view of well-being, it follows from these facts about R_i and R_j that individual i is better off with the first bundle, and individual j with the second. Thus if $u(.)$ is an accurate measure of well-being, it should assign a higher number to $(c, \mathbf{a})$ than $(c^+, \mathbf{a}^+)$ when held by i, and a lower number when held by j. We therefore need preferences *themselves* to be an argument for the well-being function: $u(.) = u(c, \mathbf{a}, R)$.

How should a $u(.)$ of the form $u(c, \mathbf{a}, R)$ be constructed? One plausible possibility, the so-called "extended preference" approach, originates with Harsanyi. The idea here is to build up $u(.)$ from the vNM functions representing the various possible preference structures (Harsanyi 1977, ch. 4, Adler 2016; but for critical discussion, see Fleurbaey 2016). That is, $u(c, \mathbf{a}, R) = \varphi^R(c, \mathbf{a})$, with $\varphi^R(.)$ one of the vNM functions representing R; $u(c, \mathbf{a}, R') = \varphi^{R'}(c, \mathbf{a})$, with $\varphi^{R'}(.)$ one of the vNM functions representing R'; and so forth. However, the vNM functions representing a given preference structure *cannot* be chosen arbitrarily.[15] Rather, we arbitrarily choose a vNM function $\varphi^R(.)$ for *one* preference structure R, and then for each vNM function $\varphi^{R'}(.)$ representing another preference structure R' identify appropriate scaling factors $s(R')$, $t(R')$, such that: $u(c, \mathbf{a}, R) = \varphi^R(c, \mathbf{a})$; $u(c, \mathbf{a}, R') = s(R')\,\varphi^{R'}(c, \mathbf{a}) + t(R')$. How to specify such scaling factors is a normative question. In effect, we are making a normative determination regarding how interpersonal comparisons among individuals with different preferences depend upon their consumption, non-consumption attributes, and preferences.[16]

Table 1[17] summarizes the different types of well-being functions that might be used as inputs to the SWF. The most general and flexible well-being function allows for variation in consumption, non-consumption attributes, and preferences. Simpler forms may be used to facilitate modelling—but the price of such simplicity is the (likely) unrealistic assumption

[15] Imagine that we define $u(.)$ using $\varphi^R(.)$ and $\varphi^{R'}(.)$. $u^+(.)$ is defined with $\varphi^R(.)$ and a different vNM representation of R', namely $f^{R'}(.)$. Then the $v^+(.)$ corresponding to $u^+(.)$ may well result in different interpersonal comparisons than $v(.)$, corresponding to $u(.)$. See examples in Adler (2016).

[16] It should be stressed that *this* scaling question—how to scale the vNM functions representing different preferences—arises whenever the "extended preferences" methodology is employed to construct a well-being function that takes account of preference heterogeneity. Both utilitarians and prioritarians must confront *this* scaling problem. Prioritarians then face the *additional* and quite distinct scaling problem described below in Part 3.2. In particular, Atkinson prioritarians will need to specify $s(.)$ and $t(.)$ for each preference and then *in addition* a zero point $(c^{zero}, \mathbf{a}^{zero}, R^{zero})$.

[17] In some policy work, including climate scholarship, consumption has been normalized to account for variation in other attributes. A detailed discussion is beyond the scope of this Article. The normalization will be in light of some preference structure establishing equivalences between changes in c and in $\mathbf{a}$. Thus, if normalized consumption is used as an input into the well-being function $u(c)$, this still assumes common preferences over $(c, \mathbf{a})$ bundles.

Table 1 Well-being Functions

Functional form of well- being measure u(.)	Embedded assumption about individual homogeneity	Construction of $u(.)$
$u(c)$	Individuals can vary with respect to consumption but have the same non-consumption attributes and preferences (see footnote 17)	Any vNM function $\varphi(c)$ representing the common preferences over consumption
$u(c, \mathbf{a})$	Individuals can vary with respect to consumption and non-consumption attributes, but have the same preferences	Any vNM function $\varphi(c, \mathbf{a})$ representing the common preferences over consumption/non-consumption bundles
$u(c, \mathbf{a}, R)$	Individuals can vary with respect to consumption, non-consumption attributes, and preferences	A matter for debate. The "extended preference" methodology chooses scaling factors for the vNM function representing each R, so that $u(c, \mathbf{a}, R) = s(R)\,\varphi^R(c, \mathbf{a}) + t(R)$

that individuals are homogeneous with respect to characteristics not incorporated in the well-being function.

3.2 Rescaling (Via a Zero Bundle or Otherwise)†

It is important to understand that *all* of the $u(.)$ functions in Table 1 are unique only up to an affine transformation, and thus yield a $v(.)$ function that is only *unique up to a common affine transformation*.[18]

A $v(.)$ unique up to a common affine transformation has a sufficient degree of interpersonal comparability for the *utilitarian* SWF. Let $v^+(.)$ be a common affine transformation of $v(.)$, i.e., there is a single positive number a and a number b s.t. $v_i^+(x) = av_i(x) + b$ for all i and x. Then the utilitarian SWF is invariant to the substitution of $v^+(.)$ for $v(.)$.[19]

However, no prioritarian SWF is invariant to common affine transformations of the well-being function $v(.)$. For every prioritarian SWF, there is some pair of outcomes, x, y and well-being function $v(.)$ s.t. $\sum_{i=1}^N g(v_i(x)) \geq \sum_{i=1}^N g(v_i(y))$ but not $\sum_{i=1}^N g(av_i(x) + b) \geq \sum_{i=1}^N g(av_i(y) + b)$, for some positive a and some b. On the informational demands of various SWFs, see Bossert and Weymark (2004) and Blackorby et al. (2005).

In effect, the information about interpersonal (and intrapersonal) comparisons of well-being levels and differences embodied in a $v(.)$ unique up to a common affine information

[18] This is clear for the well-being functions $u(c)$ and $u(c, \mathbf{a})$, as vNM functions are unique only up to an affine transformation. The generalized $u(c, \mathbf{a}, R)$ is also unique only up to an affine transformation (see Adler 2016). Note that $u(.)$ and $au(.)+b$, a positive, yield the very same ranking of $(c, \mathbf{a}, R)$ bundles and differences between them.

[19] Recall that we are assuming a fixed population. The variable-population case, discussed in Part 6, raises different issues.

is not *sufficient* well-being information for prioritarianism. More information needs to be generated. How?

One piece of additional information concerns well-being *ratios*. Well-being function $v^+(.)$ is a common *ratio* transformation of $v(.)$ iff there exists a positive number a such that $v_i^+(x) = av_i(x)$ for all i and x. The only prioritarian SWFs that are invariant to a common *ratio* transformation of well-being numbers are SWFs within the Atkinson subfamily (Adler 2012, ch. 5).

This feature of the Atkinson SWF—ratio-rescaling invariance—indeed provides a powerful normative argument in favor of that SWF, as opposed to some non-Atkinson prioritarian SWF. Welfarists should find it problematic that an SWF $w(.)$ violates ratio-rescaling invariance. Consider any case in which $w(v(x)) \geq w(v(y))$ but not $w(v^+(x)) \geq w(v^+(y))$, with $v^+(.) = av(.)$ and a positive. The two well-being functions $v^+(.)$ and $v(.)$ are *identical* in their inter- and intrapersonal comparisons of well-being levels, differences, *and* ratios. Nonetheless, SWF $w(.)$ differentiates between the two. This SWF behaves as if there is extra information about well-being—beyond information about well-being levels, differences, and ratios—concerning which the two functions disagree. But it is hard to see what such extra information could consist in.

In order to arrive at a $v(.)$ unique up to a common ratio transformation, we proceed as follows. We start with our $u(.)$ function from Table 1 above and then specify a "zero bundle." The choice of this zero bundle determines ratios: the ratio between any two bundles is just the ratio of their differences from the zero bundle. We now define a new well-being function $u^*(.)$, equaling $u(.)$ minus the $u(.)$ value of the zero bundle. Note that $u^*(.)$ and $u(.)$ agree in their assignments of well-being levels and differences, but $u^*(.)$ *also* embodies the ratio information arising from the choice of the zero bundle.[20]

The well-being measure $u^*(.)$ is unique up to a ratio transformation, and $v^*(.)$ corresponding to $u^*(.)$ is unique up to a common ratio transformation.[21] And the Atkinson SWF can be defined as follows: $w(x) = (1-\gamma)^{-1} \sum_{i=1}^{N} v_i{}^*(x)^{1-\gamma}$.

For example, if $u(.)$ has the most general form in Table 1 above, then the zero bundle is some combination of consumption, non-consumption attributes, and preferences $(c^{zero}, \mathbf{a}^{zero}, R^{zero})$. $u^*(.)$ is defined as follows: $u^*(c, \mathbf{a}, R) = u(c, \mathbf{a}, R) - u(c^{zero}, \mathbf{a}^{zero}, R^{zero})$. Similarly, in the case where individuals vary in both consumption and non-consumption attributes, but have common preferences, $u^*(.)$ is defined by specifying a zero (c, **a**) bundle: $u^*(c, \mathbf{a}) = u(c, \mathbf{a}) - u(c^{zero}, \mathbf{a}^{zero})$. And in the simplest case with well-being just a function of consumption, $u^*(.)$ is specified via a zero level of consumption. $u^*(c) = u(c) - u(c^{zero})$. Note that, in all of these cases, the $u^*(.)$ function by construction takes the value 0 at the zero bundle.

How should the zero bundle be selected? This is a critical normative question for Atkinson prioritarians. Two important features of the zero bundle should be emphasized.

---

[20] Let B and B^+ be bundles, whether of the form (c), $(c, \mathbf{a})$, or $(c, \mathbf{a}, R)$. Specifying a zero bundle B^{zero} *defines* well-being ratios; the ratio between two bundles is just the ratio of their differences from the zero bundle. That is, the well-being ratio between B and B^+ is just $[u(B) - u(B^{zero})]/[u(B^+) - u(B^{zero})]$. Note that inserting a different zero bundle B' in this formula will lead to different ratios, if $u(B^{zero}) \neq u(B')$.
Now consider the function $u^*(.)$ defined as follows: $u^*(B)$ for any bundle equals $u(B) - u(B^{zero})$. Note that $u^*(.)$ is an affine transformation of $u(.)$ and thus contains exactly the same information as $u(.)$ about the well-being levels of bundles and differences between them. In addition, $u^*(B)/u^*(B^+)$ equals the ratio between the bundles as defined by the choice of B^{zero} as zero point. See Adler (2012, ch. 3, 5) on these issues.

[21] Consider $u^*(.)$ as defined in the footnote immediately above, such that $u^*(B^{zero}) = 0$. If another $u^{**}(.)$ also represents the well-being difference and level information in $u^*(.)$, then $u^{**}(.) = au^*(.) + b$, a positive; and if $u^{**}(.)$ also implies the same ratios as $u^*(.)$, then $u^{**}(B^{zero}) = 0$ and $b = 0$.

3.2.1 The Point of Absolute Priority

The marginal moral impact of well-being, for a given SWF, is the derivative of the SWF with respect to well-being—the change in the value of the SWF per unit of well-being. Consider a given bundle B—which might take the general form $(c, \mathbf{a}, R)$ or the more restricted forms $(c, \mathbf{a})$ or (c). In the case of the Atkinson SWF, with $u^*(.)$ as the well-being function, the marginal moral impact of well-being at bundle B is just $u^*(B)^{-\gamma}$.[22] Note that the marginal moral impact of well-being is finite for any bundle better than the zero bundle (any bundle with a u^* value > 0), but approaches infinity as the bundle gets close to the zero bundle (as u^* approaches 0). Thus the *ratio* of the marginal moral impact of well-being at two bundles (a kind of marginal rate of substitution) also approaches infinity as one of the two approaches the zero bundle. (This is true, observe, *regardless* of the choice of inequality aversion parameter γ.) That is to say, the zero bundle is such that: if we consider an increment Δu^* to well-being which is smaller and smaller, and a badly off individual whose bundle is closer and closer to the zero bundle, the ratio between the moral good done by conferring that increment upon the badly off individual and conferring that same increment upon someone with a better bundle becomes indefinitely large.[23]

We can make similar observations about the marginal moral impact of *consumption*. The marginal moral impact of consumption for the Atkinson SWF—the derivative with respect to *consumption*— is just $u^*(B)^{-\gamma}\frac{\partial u^*(B)}{\partial c}$, which again is finite at any bundle better than the zero bundle, but approaches infinity as the bundle approaches the zero bundle.[24] Thus if we consider an increment Δc to *consumption* which is smaller and smaller, and a badly off individual whose bundle is closer and closer to the zero bundle, the ratio between the moral good done by conferring that increment upon the badly off individual and conferring that same increment upon someone with a better bundle becomes indefinitely large.

In both of these senses, the zero bundle is a kind of point of absolute priority. The Atkinson SWF does not in general give absolute priority to worse off individuals (see Part 4), but it does give absolute priority to someone holding the zero bundle.

3.2.2 Negative Well-Being

The Atkinson SWF cannot be used to rank outcomes in which any individuals are assigned *negative* well-being numbers. An outcome x in which $v_i^*(x) < 0$ for some individual i lies outside the domain of the Atkinson SWF. Why? Depending on the value of γ, the Atkinson SWF is either undefined at outcome x or, in the neighborhood of x, is not prioritarian (Adler 2012, p. 391).[25]

[22] With $w(v^*) = \frac{1}{1-\gamma}\sum_{i=1}^{N}(v_i^*)^{1-\gamma}$, $\frac{\partial w}{\partial v_i^*} = (v_i^*)^{-\gamma}$. If i holds bundle B, v_i^* is just $u^*(B)$.

[23] For a person at a given well-being level u^*, the moral benefit according to the Atkinson SWF of adding an increment Δu^* is $(1-\gamma)^{-1}[(u^* + \Delta u^*)^{1-\gamma} - u^{*1-\gamma}]$. Note that, with $\gamma \geq 1$, this expression is not defined at the zero bundle itself, where $u^* = 0$. However, for all values of $\gamma > 0$, we can define the ratio between the marginal moral impact of well-being at the zero bundle, and the marginal moral impact at some better bundle B such that $u^*(B) = L$, as follows: $\lim_{u^*\to 0}\lim_{\Delta u^*\to 0}\frac{(u^*+\Delta u^*)^{1-\gamma}-u^{*1-\gamma}}{(L+\Delta u^*)^{1-\gamma}-L^{1-\gamma}}$. For any $u^*(B) = L > 0$, this limit is infinite.

[24] We assume that the marginal utility of consumption, the second term in this formula, is finite and positive at every bundle except, perhaps, the zero bundle.

[25] This is because the function $(1-\gamma)^{-1}(u^*)^{1-\gamma}$ is either undefined or, if defined, not both strictly increasing and strictly concave with negative values of u^* in the domain of the function. In the case of $\gamma \geq 1$, the above function is also not defined if $u^* = 0$; if so, the zero bundle itself cannot belong to any of the outcomes being ranked.

Table 2 The Relevance of the Zero Bundle

	Atkinson SWF	Utilitarian SWF
Marginal moral impact of consumption at bundle c_1	$(\log c_1 - \log c^{zero})^{-\gamma} \times (1/c_1)$	$1/c_1$
Marginal moral impact of consumption at bundle c_2	$(\log c_2 - \log c^{zero})^{-\gamma} \times (1/c_2)$	$1/c_2$
Ratio of impacts	$\left(\frac{\log c_2 - \log c^{zero}}{\log c_1 - \log c^{zero}}\right)^{\gamma} \times \frac{c_2}{c_1}$	$\frac{c_2}{c_1}$
Ratio of marginal moral impact of consumption at c^{zero} to impact at any higher consumption level c	$\left(\frac{\log c - \log c^{zero}}{0}\right)^{\gamma} \times \frac{c}{c^{zero}}$	$\frac{c}{c^{zero}}$

These considerations suggest that the zero bundle for a given outcome set should be set at or below the lowest possible bundle for that set of outcomes. In particular, they suggest that c^{zero} should be at or below the subsistence level of consumption—the lowest level of consumption at which someone can remain alive.[26] But this is a topic for normative debate: the "lowest possible" approach to setting the zero bundle can be contested.[27]

Table 2 illustrates the relevance of the zero bundle for purposes of the Atkinson SWF, and the sense in which the zero bundle is the point of absolute priority. Assume a logarithmic consumption-only well-being function: $u(.) = u(c) = \log c$. Then $u^*(.)$ is specified via the choice of $c^{zero} : u^*(c) = u(c) - u(c^{zero})$, with $c^{zero} > 0$. Consider now two individuals with different consumption bundles above c^{zero}, individual 1 with consumption c_1 and individual 2 with consumption c_2. Table 2 calculates the marginal moral impact of consumption at the levels of these two individuals, and the *ratio* of these marginal moral impacts. It does this calculation both for the Atkinson SWF, and for the utilitarian SWF. What can be seen is that the ratio of the Atkinson marginal moral impact of consumption at level c_1 versus c_2 *depends* on the choice of c^{zero}. By contrast, the ratio of the utilitarian marginal moral impacts is independent of the choice of c^{zero}.[28]

The table also calculates the ratio of the marginal moral impact of consumption at c^{zero} itself, to the marginal moral impact at any higher level c. This ratio is infinite for the Atkinson SWF for $c^{zero} > 0$ (and undefined if $c^{zero} = 0$), but finite for the utilitarian SWF except if $c^{zero} = 0$.

This discussion about the absolute priority point relates to Weitzman (2009)'s "dismal theorem." This theorem is obtained under a utilitarian SWF and a CRRA wellbeing function (see Millner 2013). It is essentially due to the fact that the marginal moral impact of consumption $\frac{\partial u(c)}{\partial c} = c^{-\alpha}$ goes to infinity at zero consumption under CRRA. By adopting a wellbeing function with $\frac{\partial u(c)}{\partial c}$ finite at zero consumption (such as under constant absolute risk aversion, CARA), or constraining consumption to be non-zero, the utilitarian can avoid the dismal theorem. Our analysis suggests that the dismal theorem may be a more entrenched problem for

26 It is sometimes analytically convenient to set $c^{zero} = zero$ consumption, but note that this precludes a CRRA well-being function $u(c) = (1-\alpha)^{-1}c^{1-\alpha}$ with the coefficient of risk aversion $\alpha \geq 1$.

27 Picking the worst possible bundle as the zero bundle might have counterintuitive implications about well-being ratios between various other bundles, or regarding the comparative marginal moral impact of well-being at them.

28 This is just because $u^*(c) = u(c) - u(c^{zero}) = \log c - \log c^{zero}$ is an affine transformation of $u(.)$ for any choice of c^{zero}, and—as discussed above—the utilitarian SWF is invariant to such transformations.

the prioritarian, since the prioritarian marginal moral impact of consumption $u^*(B)^{-\gamma}\frac{\partial u^*(B)}{\partial c}$ become infinite at the zero bundle as long as $\frac{\partial u^*(B)}{\partial c}$ is *positive*, and regardless of whether c^{zero} is set as zero or some higher level.

This section has focused on prioritarian SWFs of the Atkinson variety. We have suggested that non-Atkinson prioritarian SWFs seem less attractive because they violate the axiom of ratio rescaling invariance. In any event, the $u(.)$ and $v(.)$ functions adequate for the utilitarian SWF are *not* adequate for any type of prioritarian SWF, since *no* prioritarian SWF is invariant to a common affine transformation of $v(.)$. The analyst using *any* type of prioritarian SWF must first rescale $u(.)$. In the case of the Atkinson SWF, this rescaling involves the choice of a zero bundle. For non-Atkinson SWFs, a different type of rescaling is required.[29]

For the remainder of the Article, we will use $u(.)$ to mean a well-being function for bundles, unique up to an affine transformation, which represents well-being levels and differences; and $v(.)$ to mean a matching vector-valued function for outcomes that is unique up to a common affine transformation. By contrast, we use $u^*(.)$ and $v^*(.)$ to mean well-being functions that have been rescaled from the $u(.)$ and $v(.)$ functions in a manner appropriate for the prioritarian SWF at issue—be it by choice of the zero bundle (for the Atkinson SWF) or in some other way (for a non-Atkinson prioritarian SWF).

4 Climate Policy Under Social Inequality Aversion†

Return to the basic formula for the prioritarian SWF: $w(x) = \sum_{i=1}^{N} g(v_i^*(x))$, with $g(.)$ strictly increasing and concave. With the Atkinson SWF, the transformation function $g(.)$ takes the form $g(v_i^*) = (1-\gamma)^{-1}(v_i^*)^{1-\gamma}$, with $\gamma > 0$.

The use of the $g(.)$ function is the mathematical device which ensures that the prioritarian SWF satisfies the Pigou–Dalton principle: the key axiomatic difference between prioritarianism and utilitarianism. But note that the use of *any* $g(.)$ strictly increasing and concave in the above formula will conform to the Pigou–Dalton principle. The choice of a particular $g(.)$—a particular value of γ, for the Atkinson SWF—presents a further normative question. The question here concerns the *degree* of priority for someone at a lower well-being level.

In the case of the Atkinson SWF, as already mentioned, the marginal moral impact of well-being at bundle B is just $u^*(B)^{-\gamma}$. Consider now two bundles B and B', such that the well-being ratio between them is $1/K$: $u^*(B) = U^*$ and $u^*(B') = KU^*$ (with K and

[29] This topic cannot be discussed at length here. Let $u(.)$ be a well-being function that represents levels and differences, and the set **U** all utility functions that are affine transformations of $u(.)$. No prioritarian SWF $\sum g(.)$ will be invariant to the use of *any* $u(.)$ in **U**. Rather, a particular subset $\mathbf{U}^* \subset \mathbf{U}$ (at the limit, a singleton subset) will be such that (1) the SWF is invariant to the use of any $u^*(.)$ in **U***, and (2) **U*** will be identified as the "right" subset by virtue of its implications for the marginal moral impact of well-being and consumption at various bundles, given $g(.)$ and given the $u^*(.)$ functions in **U***.
For a given arbitrary $u(.)$ in **U**, there will be some affine transformation(s) of $u(.)$ that belongs to **U***: $u^*(.) = cu(.) + d$, with c taking a specific positive value (or range of values), and d some specific value (or range of values). In the particular case of the Atkinson SWF, because **U*** consists of every ratio transformation of some $u^+(.)$, it follows that if some well-being function in **U*** assigns the number zero to some bundle B^{zero}, then every other one does; and that one rescaling which transforms an arbitrary $u(.)$ into a member of **U*** is $u(.) - d$, where $d = u(B^{zero})$.
However, this particular strategy for rescaling an arbitrary $u(.)$ does not generalize to non-Atkinson prioritarian SWFs. Consider the negative exponential SWF, $w(x) = \sum_{i=1}^{N} - \exp(-kv_i^*(x))$, $k > 0$. **U*** here consists of some $u^+(.)$ and every other well-being function equaling $u^+(.) + b$. Note that there is *not* some B^{zero} such that $u^*(B^{zero}) = 0$ for every $u^*(.)$ in **U***. Moreover, for an arbitrary $u(.)$ in **U**, it is *not* true that there is necessarily some d such that $u(.)$ can be rescaled into some $u^*(.)$ in **U*** by the rescaling $u(.) - d$ (indeed this will never be true if $u(.)$ is not already in **U***).

$U^* > 0$). Then the ratio of marginal moral well-being impacts is just K^γ. That is to say, giving a small increment of well-being to an individual at a given well-being level, U^*, has an ethical impact K^y times greater than giving that same increment to an individual K times better off at KU^*.

Holding constant the well-being ratio K between the better- and worse off individual, K^y increases as γ does—the degree of priority for the worse-off one increases. (That degree of priority is insensitive to γ—it becomes infinite—only when the well-being of the worse-off one is at the point of absolute priority $U^* = 0$). Adler (2012, ch. 5) describes normative thought experiments that the ethical deliberator might use to fix γ. For empirical works on social inequality aversion, see for instance Carlsson et al. (2005), and Gaertner and Schokkaert (2012).

It should be stressed that γ (and, more generally, $g(.)$) is an *ethical* parameter embodying aversion to inequality, and should not be conflated with risk aversion parameter α of the CRRA consumption-utility function—$u(c) = (1-\alpha)^{-1}c^{1-\alpha}$—which captures individuals' self-interested preferences over consumption gambles (Kaplow 2010; Kaplow and Weisbach 2011).

In this section, we illustrate, with reference to climate change, why the choice of $g(.)$ is a crucial issue for the prioritarian. We will consider some simple implications of $g(.)$ for climate change under three headings: the "social cost of carbon," the "social discount rate," and optimal mitigation.

4.1 The Social Cost of Carbon

The "social cost of carbon" (SCC) is the damage to social welfare per unit of emissions, expressed in terms of the equivalent money (consumption) loss. The SCC is the critical construct for incorporating climate impacts into cost-benefit analysis (CBA), and indeed now plays a central role in governmental CBA in the U.S. (Greenstone et al. 2013; Tol 2011; van den Bergh and Botzen 2014).

The SCC depends on the date of emissions; on the trajectory of consumption and other attributes that are arguments in the well-being function; and on those features of the climate system and human society that are variables for the "damage function" whereby emissions cause damage to the well-being attributes. The SCC depends, further, on the "numeraire" date at which the consumption equivalent is being calculated, and upon the incidence of consumption costs. These points are common in the literature. Less well recognized is that the SCC *also* depends upon the SWF. The utilitarian SCC, as calculated by the literature, should be distinguished from a prioritarian SCC—with that latter value in turn depending on the choice of $g(.)$.

To make the illustration simple, we ignore intragenerational heterogeneity in consumption, non-consumption attributes, or preferences. We assume that individual well-being depends upon consumption, a single non-consumption attribute h (for short, "health"), and preferences over (c, h) bundles, so that $u(.) = u(c, h, R)$, and $u^*(.) = u(c, h, R) - u(c^{zero}, h^{zero}, R^{zero})$, or some other appropriate rescaling if a non-Atkinson prioritarian SWF is being used. Because the utilitarian SWF and SCC are invariant to the substitution of $u^*(.)$ for $u(.)$,[30] while the prioritarian SWF and SCC are not, we use $u^*(.)$ in both the utilitarian and prioritarian calculations. This will make the comparison of the prioritarian and utilitarian cases more transparent.

[30] Since $u^*(.)$ is always an affine transformation of $u(.)$, see footnote 29, this must hold true.

Carbon emissions damage future consumption and health, but do not change preferences. An "outcome" such as x not only describes individuals' welfare attributes, but also emissions amounts and the geophysical and societal features that figure in the damage function. (Thus, for a given outcome, we can predict what damage will occur with an incremental unit of emissions.) The SCC, at outcome x, for a given emissions date and numeraire date, is the ratio of the marginal moral impact of emissions (at the emissions date) to the marginal moral impact of consumption (at the numeraire date)—as calculated by a given SWF.

Let c_t^x, h_t^x, and R_t^x represent the time t consumption, health and preference, respectively in a given outcome x. Let e_l be carbon emissions at time l. Let $\frac{\partial c_t^x}{\partial e_l^x}$ denote the damage to consumption at time t, per unit of emissions, that would result in outcome x from emissions at time l—and similarly $\frac{\partial h_t^x}{\partial e_l^x}$ denotes health damage per unit of emissions. Let $SCC^{util}(x, l, l^*)$ denote the utilitarian SCC around x, for emissions at l and with numeraire at l^*. Without loss of generality, we set the numeraire to time 1, the present.

$$SCC^{util}(x, l, 1) = \frac{\sum_{t=1}^{T\max} N_t \left[\frac{\partial u^*}{\partial c}\left(c_t^x, h_t^x, R_t^x\right) \frac{\partial c_t^x}{\partial e_l^x} + \frac{\partial u^*}{\partial h}\left(c_t^x, h_t^x, R_t^x\right) \frac{\partial h_t^x}{\partial e_l^x} \right]}{\frac{\partial u^*}{\partial c}\left(c_1^x, h_1^x, R_1^x\right)}$$

Note that the numerator here expresses the effect on consumption and health at all future dates of a small change in emissions at time l, with each such change then translated into a well-being change by multiplying by the marginal utility of consumption or health at that time.

The *prioritarian* social cost of carbon is defined similarly—except that the change in social welfare arising from a change in consumption Δc_t or a change in health Δh_t is not the corresponding well-being change, i.e., $\frac{\partial u^*}{\partial c}(c_t^x, h_t^x, R_t^x)\Delta c_t$ or $\frac{\partial u^*}{\partial h}(c_t^x, h_t^x, R_t^x)\Delta h_t$, but the change in *transformed* well-being, i.e., $g'(u^*(c_t^x, h_t^x, R_t^x))\frac{\partial u^*}{\partial c}(c_t^x, h_t^x, R_t^x)\Delta c_t$ or $g'(u^*(c_t^x, h_t^x, R_t^x))\frac{\partial u^*}{\partial h}(c_t^x, h_t^x, R_t^x)\Delta h_t$.

$$\begin{aligned} &SCC^{prior}(x, l, 1) \\ &= \frac{\sum_{t=1}^{T\max} N_t g'\left(u^*\left(c_t^x, h_t^x, R_t^x\right)\right) \left[\frac{\partial u^*}{\partial c}\left(c_t^x, h_t^x, R_t^x\right) \frac{\partial c_t^x}{\partial e_l^x} + \frac{\partial u^*}{\partial h}\left(c_t^x, h_t^x, R_t^x\right) \frac{\partial h_t^x}{\partial e_l^x} \right]}{g'\left(u^*\left(c_1^x, h_1^x, R_1^x\right)\right) \frac{\partial u^*}{\partial c}\left(c_1^x, h_1^x, R_1^x\right)} \end{aligned}$$

The prioritarian SCC can clearly differ from the utilitarian SCC; and different specifications of $g(.)$ will yield different prioritarian values.

We can say more. Assume that future generations are better off than the present generation. Then it can be seen that the prioritarian SCC will be *less than* the utilitarian SCC. Why? By the strict concavity of $g(.)$, $u^*(c_t, h_t, R_t) > u^*(c_1, h_1, R_1)$ for $t > 1$ implies that $g'(u^*(c_t, h_t, R_t)) < g'(u^*(c_1, h_1, R_1))$ and thus $g'(u^*(c_t, h_t, R_t))/g'(u^*(c_1, h_1, R_1)) < 1$ for all $t > 1$.[31] Intuitively, with a future better off than the present, prioritarians place less moral weight than utilitarians on the loss of well-being caused by emissions, as compared with the moral weight of welfare loss caused by reduced present consumption.

4.2 The Social Discount Rate

The social discount rate can be introduced using the following simple problem (Gollier 2013), reverting to the simplest form of well-being $u(c)$, assumed to be increasing in consumption.

[31] $g'(.)$ is always positive because $g(.)$ is strictly increasing.

Suppose that a safe project costs a small amount Δc today, in period 1, and returns $\Delta c(1+r)^t$ in period $t+1$, $t \geq 1$, with c_1 the (exogenous) consumption today and c_{t+1} the consumption in period $t+1$. If we use the utilitarian SWF to evaluate the project, the project's marginal benefit is $-u'(c_1)\Delta c + u'(c_{t+1})\Delta c(1+r)^t$, with primes denoting first derivatives. This value is positive, i.e., the project is approved by the utilitarian SWF, iff the rate of return r of the project is greater than the utilitarian social discount rate r^{util}, defined as follows:

$$r^{util} = \left[\frac{u'(c_1)}{u'(c_{t+1})}\right]^{1/t} - 1.$$

Consider now a prioritarian SWF. We have $u^*(c) = u(c) - u(c^{zero})$ in the case of the Atkinson SWF (c^{zero} less than consumption in the two periods), or some other appropriate rescaling for a non-Atkinson SWF. The *prioritarian* marginal net benefit of the project is[32] $-u'(c_1)g'(u^*(c_1))\Delta c + u'(c_{t+1})g'(u^*(c_{t+1}))\Delta c(1+r)^t$. *This* value is positive iff the rate of return r of the project is greater than the *prioritarian* social discount rate r^{prior}, defined as follows:

$$r^{prior} = \left[\frac{g'(u^*(c_1))u'(c_1)}{g'(u^*(c_{t+1}))u'(c_{t+1})}\right]^{1/t} - 1$$

Note that $r^{prior} > r^{util}$ iff $c_{t+1} > c_1$,[33] i.e., there is positive growth between the two periods. The intuition for this result is similar to that in the previous section. Under positive (respectively negative) growth, a project which transfers wealth into the future is less (respectively more) valuable under prioritarianism than under utilitarianism.

Assume now a CRRA well-being function $u(c) = (1-\alpha)^{-1}c^{1-\alpha}$ together with $c_{t+1} = (1+r_e)^t c_1$, where r_e is the rate of growth of the economy. Observe first that the utilitarian social discount rate r^{util} equals $(1+r_e)^\alpha - 1$, which can be approximated by αr_e.[34] This is simply the well-known Ramsey rule with a zero rate of pure time preference. Note that the utilitarian social discount rate is independent from t; the term structure is "flat" in this particular case.

What does the Ramsey rule look like under prioritarianism? Observe that the formula above for r^{prior} includes the term $\left[\frac{g'(u^*(c_1))}{g'(u^*(c_{t+1}))}\right]^{1/t}$, which indicates that the term structure cannot generically be flat without further assumptions about the type of prioritarian SWF. We therefore now assume an Atkinson SWF with inequality aversion parameter $\gamma > 0$. With this assumption and the fact that $u^*(c) = u(c) - u(c^{zero})$, we obtain a *prioritarian* Ramsey formula. $r^{prior} \approx r_e\left[\alpha + \gamma\frac{1-\alpha}{1-(c_1/c^{zero})^{\alpha-1}}\right]$.[35]

This prioritarian Ramsey formula shows that r^{prior}—under the assumption of CRRA well-being and an Atkinson SWF—is independent from t, as with r^{util}. The term $r_e\gamma\frac{1-\alpha}{1-(c_1/c^{zero})^{\alpha-1}}$ captures the effect on the social discount rate of shifting from utilitarianism to prioritarianism,

[32] Because $u^*(.) = au(.) + b$, a positive, the first derivative of $u^*(.)$ is just $au'(.)$, and we have (without affecting the formula for the prioritarian discount rate immediately below) divided both sides by a.

[33] $g'(u^*(c_1)) > g'(u^*(c_{t+1}))$ iff $c_{t+1} > c_1$. Note that $u^*(.)$ is increasing in consumption since $u(.)$ is.

[34] If $f(x) = x^\alpha$, $f(1+\Delta x) \approx f(1) + f'(1)\Delta x = 1 + \alpha\Delta x$.

[35] With $\alpha = 1$, we have $r^{prior} \approx r_e\left[\alpha + \gamma\frac{1}{\log(c_1/c^{zero})}\right]$. To see this, just apply L'Hospital's rule, $\lim_{\alpha\to 1}\frac{f_1(\alpha)}{f_2(\alpha)} = \frac{f_1'(1)}{f_2'(1)}$, with $f_1(\alpha) = 1-\alpha$ and $f_2(\alpha) = 1-(c_1/c^{zero})^{\alpha-1}$.

since $r^{util} \approx r_e\alpha$ and thus $r^{prior} \approx r^{util} + r_e\gamma \frac{1-\alpha}{1-(c_1/c^{zero})^{\alpha-1}}$. Note that this term is positive iff there is positive growth ($r_e > 0$), since $c_1 > c^{zero}$.[36]

Moreover, observe that with positive growth, the term $r_e\gamma \frac{1-\alpha}{1-(c_1/c^{zero})^{\alpha-1}}$ and thus r^{prior} is increasing in both γ and c^{zero}. It becomes arbitrarily large as γ does, and as c^{zero} approaches c_1. This last observation reflects the status of the zero bundle as the point of absolute moral priority within Atkinson prioritarianism. Finally, with risk neutrality ($\alpha = 0$) and positive growth, r^{util} is zero but r^{prior} is positive, reflecting the effect of inequality aversion.

To be sure, the analysis here (as that of the SCC above) uses a simple model without intratemporal heterogeneity. We now turn to the problem of optimal mitigation and, as part of that analysis, allow for such heterogeneity.

4.3 Optimal Mitigation

Integrated Assessment Models calculate optimal policies (with respect to emissions, investment, and other choice variables) by identifying the policies that produce an optimal time path of consumption and, ideally, other welfare-relevant attributes. Here, we make some initial observations about optimal consumption with a simple model (Dasgupta 2008). We assume an Atkinson SWF and CRRA consumption-based well-being: $u(c) = (1-\alpha)^{-1}c^{1-\alpha}$, and $u^*(c) = u(c) - u(c^{zero})$. To illustrate some basic insights and make the optimization tractable, we set $c^{zero} = 0$, which in turn constrains the coefficient of risk aversion α to be less than 1.

Assume that there are two time periods. The population size N is identical in both periods. There is a fixed total of potential output K in the first period. Any or all of it can be consumed in the first period, or invested at some fixed positive rate r. ("Investment" is meant to capture both physical investment, and the increment to future consumption that comes from abating emissions by declining to produce some part of potential output K.) Thus if c_1 and c_2 are total population consumption in the respective periods, the social planner operates under the "budget" constraint that $c_1 + c_2/(1 + r) = K$.

Consider first the case of intragenerational homogeneity, where c_1 and c_2 are shared equally among the populations in each period. The utilitarian planner maximizes $Nu(c_1/N)+ Nu(c_2/N)$, subject to the budget constraint. It is straightforward to show that the optimal utilitarian consumption in each period (denoted with a "+") is as follows: $c_2^+ = c_1^+(1+r)^{1/\alpha}$. Consumption is always greater in the future, increasing with r and diminishing with α. The larger r is, the more second period consumption will increase for a unit decrease in the first period. The larger α is, the smaller the well-being gain to be had from shifting consumption to the second period from the first.

What happens if the planner becomes prioritarian? She now maximizes $N(1 - \gamma)^{-1}(u(c_1/N)^{1-\gamma} + u(c_2/N)^{1-\gamma})$, subject to the budget constraint. The prioritarian optimum (denoted with "++") is as follows: $c_2^{++} = c_1^{++}(1+r)^{1/[\alpha+\gamma(1-\alpha)]}$. Note that with $\alpha < 1$ and $\gamma > 0$, the exponent in this formula is less than the exponent ($1/\alpha$) in the utilitarian formula. Thus c_2^{++} is less than c_2^+ and c_1^{++} is greater than c_1^+. Again, the prioritarian is less impressed than the utilitarian by the consumption gains to be had by shifting consumption from a poorer present to a richer future. As the coefficient of inequality aversion γ approaches infinity, the optimal prioritarian allocation approaches equal consumption in the two periods.

Now let us turn to the possibility of intragenerational variation, and for simplicity two individuals in the population in each period. Assume that in period 1 consumption is split

[36] The model here allows for $r_e < 0$, but only if $c_{t+1} = c_1(1+r_e)^t > c^{zero}$, and the approximating formula just stated should be used with this restriction on r_e in mind.

equally; in period 2, it may be split unequally, with the less well-off individual receiving only fraction π of c_2, with $\pi \le 1/2$, and the other individual receiving fraction $(1-\pi) \ge 1/2$. For simplicity, now, assume that well-being is linear in consumption ($\alpha = 0$). The utilitarian planner, optimizing under the budget constraint (and given linear well-being), finds that $c_1^+ = 0$ and $c_2^+ = K(1+r)$. The prioritarian planner's decision is much more subtle. It turns out that:

$$c_2^{++} = \frac{c_1^{++}}{2}(1+r)^{1/\gamma}[\pi^{1-\gamma} + (1-\pi)^{1-\gamma}]^{1/\gamma}$$

This is an interesting formula. There are subtle interactions between the growth rate r of invested output, the coefficient of inequality aversion γ, and the degree of future inequality π that emerge even in the very simple model now on the table. We can make a couple of initial observations about the prioritarian optimum identified by this formula. (1) If future consumption is maximally unequal ($\pi = 0$),[37] then the prioritarian planner may choose to have lower total (and average) future consumption than total (and average) present consumption. This happens if $(1+r)^{1/\gamma} < 2$. Moreover, with $\pi = 0$ she saves less and less for the future (by physical investment or abatement) the larger the coefficient of inequality aversion. (2) With the degree of future inequality now allowed to vary, observe that the ratio between c_2^{++} and c_1^{++} is *increasing* in π for $\gamma < 1$; invariant to that ratio for $\gamma = 1$ (with the Atkinson function becoming the logarithm); and *decreasing* in π for $\gamma > 1$. For the prioritarian, decreases in future inequality may either increase or decrease current consumption depending on the level of inequality aversion. This latter result shows both the sensitivity of policy choice to γ, and the importance (for the prioritarian) of using a model that allows for within- as well as across-generation inequality.[38]

5 Social Choice Under Risk

By "risk," we mean that the decisionmaker can attach a single probability to each outcome (be this an epistemic probability or an "objective" relative frequency). Given space constraints, we cannot discuss the topic of utilitarian and prioritarian decisionmaking without such probabilities (the setting for the literature on so-called "ambiguity"; see for instance Millner et al. 2013).

We introduce "risk" in the classical manner. Let s be a possible state of the world; S the set of all states; π_s the probability of s; and x_a^s the outcome of a in state s. The "ex post" social planner (be she utilitarian or prioritarian) behaves consistently with expected utility theory *at the level of social welfare*. That is, the ex post utilitarian social planner ranks choices in accordance with the following formula, with $W^{util}(a)$ the expected utilitarian social welfare of choice a.

$W^{util}(a) = \sum_{s \in S} \pi_s F(\sum_{i=1}^N v_i(x_a^s))$, with $v_i(x_a^s)$ the well-being of individual i in the outcome that results from state s given action a.

37 This part of the analysis assumes $\gamma < 1$.

38 Interestingly, note that the prioritarian SWF assumes *both* within- and across-generation inequality aversion. However, we could consider as an alternative only one or the other form of inequality aversion. Indeed, it seems plausible that a decisionmaker might care *only* about differences in well-being across generations, or *only* about differences in well-being across regions. This provides interesting settings for a comparative statics analysis of different forms of inequality aversion.

$F(.)$ here is any (strictly) increasing function. Most often, in discussions of utilitarianism under uncertainty, F is assumed to be the identity function. But it is perfectly consistent with the axioms of expected utility theory for F to be any increasing function. Just as an individual (consistent with expected utility theory) can rank consumption gambles according to the expected value of any increasing function of consumption, so the utilitarian social planner (consistent with expected utility theory) can rank social-welfare gambles according to the expected value of any increasing function of the sum of individual well-being.

Similarly, the ex post prioritarian social planner ranks choices in accordance with the expected value of some (strictly) increasing function of the sum of *transformed* individual well-being numbers. She orders choices according to their prioritarian expected social welfare, $W^{prior}(a)$, calculated as follows.

$$W^{prior}(a) = \sum_{s \in S} \pi_s H\left(\sum_{i=1}^{N} g(v_i^*(x_a^s))\right),$$

with $g(.)$ the prioritarian transformation function (strictly increasing and concave, perhaps Atkinson) and $H(.)$ any strictly increasing function.

What are the fresh normative puzzles that the prioritarian must confront, here—puzzles above and beyond those that arise in the case of known outcomes? The first is deciding whether to use the "ex post" formula just articulated, or instead to follow an "ex ante" approach. The second (if she uses the "ex post" formula) is identifying the function H. The latter puzzle is also one for the utilitarian social planner, while the first is specific to prioritarianism.

5.1 Ex Ante Versus Ex Post Prioritarianism†

The "ex ante" prioritarian does not rank choices according to $W^{prior}(a)$. Instead, she sees each choice as a vector of individual expected well-being numbers, and sums these individual expected well-beings transformed by the $g(.)$ function. That is, she assigns each choice a value $W^{prior-exante}(a)$ equaling: $\sum_{i=1}^{N} g(\sum_{s \in S} \pi_s v_i^*(x_a^s))$. See Adler (2012, ch. 7), generally discussing the difference between ex post and ex ante prioritarianism.

Why is the choice between the ex ante and ex post approaches specific to prioritarianism? Can't one also define an ex ante utilitarian approach, whereby choices are ranked with this formula: $W^{util-exante}(a) = \sum_{i=1}^{N} \sum_{s \in S} \pi_s v_i(x_a^s)$? Note, though, that this formula is mathematically equivalent to ex *post* utilitarianism setting F as the identity function. By contrast, ex ante *prioritarianism* is not equivalent (in its ranking of choices) to *any* version of ex post prioritarianism, with any H.

To see why ex post and ex ante prioritarianism diverge, consider the following case.

	Action a		Expected well-being	Action b		Expected well-being
	State s $v^*(x)$	State s^+ $v^*(y)$		State s $v^*(z)$	State s^+ $v^*(w)$	
Individual A	20	60	40	10	80	45
Individual B	80	40	60	90	20	55

Explanation Each state has probability 1/2. In states s and s^+, action a produces outcomes x and y, respectively; while action b produces outcomes z and w, respectively. The columns show the realized well-being of each of two individuals with these outcomes, and their expected well-being for each action.

Note that action a can be reached from action b by a state-by-state Pigou–Dalton transfer. Thus the ex post prioritarian (whatever her H function) will prefer a. But action b can be reached from a by a Pigou–Dalton transfer in *expected well-being*. Thus the ex ante prioritarian necessarily prefers b.

The choice between ex post and ex ante prioritarianism clearly affects how the prioritarian social planner evaluates climate policies. First, consider how the "prioritarian social cost of carbon," discussed above in Part 4.1, might be extrapolated to conditions of risk. As there, we simplify by ignoring intragenerational variation. We also set H here to be the identity function. Finally, we assume that time 1 attributes are known rather than risky.

The symbols are the same as in the discussion of the SCC above, or new symbols to handle risk. To avoid clutter, we drop an explicit reference to states and instead introduce the symbol p_a^x, indicating the probability of outcome x given action a.[39] The ex post prioritarian SCC (as a function of the choice a, date of emissions l, and numeraire time 1) is as follows:

$$SCC^{EPP}(a, l, 1) = \frac{\sum_{t=1}^{T\max} N_t \sum_x p_a^x \left[g'\left(u^*\left(c_t^x, h_t^x, R_t^x\right)\right) \left[\frac{\partial u^*}{\partial c}\left(c_t^x, h_t^x, R_t^x\right) \frac{\partial c_t^x}{\partial e_l^x} + \frac{\partial u^*}{\partial h}\left(c_t^x, h_t^x, R_t^x\right) \frac{\partial h_t^x}{\partial e_l^x} \right]\right]}{g'\left(u^*\left(c_1, h_1, R_1\right)\right) \frac{\partial u^*}{\partial c}\left(c_1, h_1, R_1\right)}$$

This formula calculates the outcome-specific change to *transformed well-being* from a unit of emissions at l (resulting from outcome-specific consumption and non-market damages and consequent outcome-specific loss in well-being), and then aggregates over outcomes. By contrast, the ex ante prioritarian SCC calculates the outcome-specific change to *well-being* from a unit of emissions at l; aggregates over outcomes; and then multiplies this term by $g'(.)$ applied to the level of expected well-being.

$$SCC^{EAP}(a, l, 1) = \frac{\sum_{t=1}^{T\max} N_t g'\left(\sum_x p_a^x u^*\left(c_t^x, h_t^x, R_t^x\right)\right) \left[\sum_x p_a^x \left[\frac{\partial u^*}{\partial c}\left(c_t^x, h_t^x, R_t^x\right) \frac{\partial c_t^x}{\partial e_l^x} + \frac{\partial u^*}{\partial h}\left(c_t^x, h_t^x, R_t^x\right) \frac{\partial h_t^x}{\partial e_l^x} \right]\right]}{g'\left(u^*\left(c_1, h_1, R_1\right)\right) \frac{\partial u^*}{\partial c}\left(c_1, h_1, R_1\right)}$$

The analysis of Adler and Treich (2014) suggests that the choice between ex ante and ex post prioritarianism may also have important implications for optimal mitigation. We study a version of the standard "cake eating" problem from the precautionary savings literature, comparing utilitarian, ex ante prioritarian, and ex post prioritarian solutions. The amount of resources available for consumption in the first period (which can be interpreted as the first generation) is known, while the amount available in the second period is risky. A social planner must decide before the risk is resolved how much to consume in the first period, and how much to save for the second period.

Under quite general conditions regarding the well-being function $u(.)$ and the prioritarian transformation function $g(.)$, we find that ex ante prioritarianism saves less for the future than utilitarianism. Why? The utilitarian maximizes the sum of expected well-being across the two periods; and under general conditions on $u(.)$, namely decreasing absolute risk aversion, this optimum is such that the expected well-being in the second period is greater than well-being in the first. The ex ante prioritarian reduces this inequality between expected well-being in the second period, and first period well-being, by shifting consumption towards the first period.

We also find that the ex *post* prioritarian shifts more resources to the future than the ex ante prioritarian, and indeed—as inequality aversion increases—does so more than the utilitarian decisionmaker. The intuition here is that ex post prioritarians focus on *state-contingent*

[39] Let y_a^s indicate the outcome y of action a in state s. Then $p_a^x = \sum_{s \in S: y_a^s = x} \pi_s$, i.e., the cumulative probability of those states that would yield x were a to be performed.

inequality in *realized* well-being between the two periods (not inequality in expected well-being), and that as inequality aversion increases the possibility of states in which the second period has low resources provides a stronger motivation to save for the future.

There is now a substantial subliterature within climate change scholarship on the topic of discounting under uncertainty (Arrow et al. 2014; Gollier 2013; Gollier and Weitzman 2010; Groom et al. 2005; Traeger 2013; Weitzman 1998, 2001). An important topic for research will be to compare the social discount rate under utilitarianism, ex ante prioritarianism, and ex post prioritarianism.

We have described the ex ante/ex post prioritarian distinction, and suggested its significance for climate change. But which approach is more attractive as a normative matter? This is a contested question among social choice theorists and philosophers. See Adler (2012, ch. 7) and Fleurbaey and Bovens (2012) for a recent presentation of the debate. Certainly strong arguments can be presented in favor of the ex post approach. In particular, the ex ante approach violates a very compelling axiom of stochastic dominance (Fleurbaey 2010).[40] It should also be noted that the ex post approach is time-consistent, while the ex ante approach is not (Hammond 1983). The climate change planner, using the ex ante approach, might plan to optimize by setting a particular time path of emissions, savings, etc., but then deviate from this plan later on. Yet, the ex ante approach respects the Pareto principle in terms of individual expected well-being, unlike the ex post approach. In this sense, the ex post approach is "paternalistic," typically leading the social planner to be more risk-averse than individuals (because $g(.)$ is concave).

5.2 What is the H Function?

Assume that we are operating within ex post prioritarianism. Recall the general formula for the prioritarian ranking of outcomes: $\sum_{i=1}^{N} g(v_i^*(x))$. Compare this to the ex post prioritarian formula for ranking risky actions: $\sum_{s \in S} \pi_s H(\sum_{i=1}^{N} g(v_i^*(x_a^s)))$. The fresh choice to be made, here, concerns the shape of the H function. Should it be the identity function? Concave? Convex? Neither? Intuitively, if normal individuals are risk averse with respect to their own consumption, social planners should be risk averse with respect to social welfare (a concave H). However, Fleurbaey has characterized an "equally distributed equivalent" approach that would actually yield a convex H (see Fleurbaey 2010; Adler et al. 2014). The prioritarian social planner becomes risk prone with respect to gambles over the sum of transformed individual well-being.

An important implication of using a nonlinear function H is that the objective function is not separable in general. This nonseparability property is often seen as a deficiency. It implies that the social value of a policy that affects only a subset of the population may depend on unaffected individuals. However, a nonseparable H may also allow us to be sensitive to certain normatively significant aspects of policy impacts. For instance, the SWF can display an aversion (or a preference) to risks that are "catastrophic" in the sense that they affect many individuals in a given state (Keeney 1980; Bommier and Zuber 2008; Fleurbaey 2010). "Catastrophe aversion" means that the social planner is not indifferent to the statistical dependence across individuals' risks. Note that the concept of catastrophe aversion might be relevant in capturing one feature of the risks caused by climate change: there is a possibility of "extreme" states of the world (the climate system) in which *many* individuals are worse off than in less extreme states, or the world's population is much smaller or even goes extinct.

[40] To understand why, recall the case presented above in which the state-dependent outcome of action a is better than that of action b in both states, and yet the ex ante approach prefers action b since it is a Pigou–Dalton transfer in expected well-being.

The problem of choosing the H function is not unique to prioritarianism, since the utilitarian under risk must make a parallel choice (identifying what we denoted above as the F function). Note, though, that there is no particular reason to believe that the most normatively attractive H is the same as the most normatively attractive F.

6 Variable Population

Climate policies can change the size of the population of interest. For the impartial decision-maker, the "population of interest" is the world's intertemporal population: prior generations now dead, everyone currently alive, and all who will be born in the future. Imagine that, in one state of nature, a "business as usual" policy will cause outcome x, whereby temperature rises so much that humanity becomes extinct in the twenty-fifth century. By contrast, in this state of nature, an aggressive abatement policy would prevent catastrophic temperature change and lead to outcome y, whereby humanity continues until a much later date and eventually becomes extinct for reasons independent of temperature rise. Let $N(x)$ and $N(y)$ denote the world's intertemporal population in, respectively, x and y. Then $N(y) \gg N(x)$.

Of course, extinction is merely the extreme example of a climatic effect on population. A high degree of warming could cause a decline in reproduction, relative to a lower-warming scenario—perhaps a massive, "catastrophic" such decline—without leading all the way to extinction. Clearly, too, climate policy can affect the size of a "population of interest" smaller than the world's intertemporal population—for example, the past, present, and future citizens of a given country. However, since the focus of this Article is ethical (impartial) decision-making, with SWFs understood as tools for guiding such decisions, we leave aside this case.

Consider, to begin, the problem of variable population size under certainty, and with a utilitarian SWF. A rich philosophical literature addresses this problem (see Parfit 1987 for a seminal text; Arrhenius 2016; Holtug 2010; Roberts 2015), as does social-choice scholarship (Broome 2004; Blackorby et al. 2005). See Millner (2013) for a climate-change application.

Two classical approaches are "total utilitarianism" and "average utilitarianism." In the case of total utilitarianism, outcomes are ranked according to total well-being. It is tempting to represent total utilitarianism by the formula $w(x) = \sum_{i=1}^{N(x)} v_i(x)$, but recall that the utilitarian $v(.)$ is arrived at from a $u(.)$ function for bundles which is unique up to an affine transformation, and so $v(.)$ is only unique up to a common affine transformation. Thus the formula just stated is not meaningful.[41] To arrive at a sensible formula, let us describe total utilitarianism (consistent with the literature) as the view which says the following: adding someone to the population increases ethical value iff the person's life is better than non-existence. Now, let $(c^{worth}, \mathbf{a}^{worth}, R^{worth})$ be a life just worth living—a life sufficiently bad that its well-being level is just equal to not existing at all. Then the total-utilitarian value of an outcome is: $w(x) = \sum_{i=1}^{N(x)} v_i(x) - u(c^{worth}, \mathbf{a}^{worth}, R^{worth})$.[42]

In the case of average utilitarianism, outcomes are ranked according to average well-being. That is, $w(x) = \frac{1}{N(x)} \sum_{i=1}^{N(x)} v_i(x)$. Average utilitarianism has serious limitations. To begin,

[41] Assume that $N(x) \neq N(y)$. Then it is possible that $\sum_{i=1}^{N(x)} v_i(x) > \sum_{i=1}^{N(y)} v_i(y)$ but $\sum_{i=1}^{N(x)} (av_i(x)+b) < \sum_{i=1}^{N(y)} (av_i(y) + b)$.

[42] Recall that, in turn, $v_i(x) = u(c_i(x), \mathbf{a}_i(x), R_i(x))$, and similarly (below) that $v_i^*(x) = u^*(c_i(x), \mathbf{a}_i(x), R_i(x))$. Except where important for purpose of exposition, we will simplify formulas here by using the v_i and v_i^* notation.
Also, for most of this part, the formulas are pitched in terms of a general well-being function, $u(c, \mathbf{a}, R)$. Revising these formulas for the case of a simpler well-being function $u(c)$ or $u(c, \mathbf{a})$ is straightforward.

it is non-separable: if x and y are identical to x^+ and y^+, respectively, except that some individual(s) who do not exist in the first two outcomes exist in the second two, each at the same well-being level in x^+ as in y^+, it is possible that $w(x) > w(y)$ but $w(x^+) < w(y^+)$. This means, for example, that an average-utilitarian decisionmaker choosing climate policies in the 21st century would need to ascertain the size and well-being levels of long-dead prior generations. Many see this as a strong objection to average utilitarianism. It should be noted, however, that once we relax the simplifying assumption of choice under certainty, utilitarianism even with a fixed population is also non-separable given a non-linear F function (as discussed above).

A more serious objection to average utilitarianism is that it violates the "negative expansion principle": bringing into being a life not worth living should never be seen as an ethical improvement. Assume that individual i does not exist in x; in y, this one individual has been added to the population, but her level of well-being is so low that it would be better for her not to exist. If average well-being in x is even lower than $v_i(y)$, then $w(y) > w(x)$.

Total utilitarianism is separable under conditions of certainty, and satisfies the "negative expansion principle," but has an unappealing feature that Parfit famously called the "repugnant conclusion." Assume that the average level of well-being in x is above the level of a life just worth living, $(c^{worth}, \mathbf{a}^{worth}, R^{worth})$, but arbitrarily close to that level. Consider an outcome y, with an arbitrarily high average well-being level. Then if $N(x) - N(y)$ is sufficiently large, $w(x) > w(y)$. Policies that add to the world's population are always ethically recommended, even if the effect of doing so is a dramatic decline in the average quality of life (to any well-being level above nonexistence), as long as the increase in population is sufficiently large.

The (arguable) flaws that characterize these two classical methodologies can be circumvented via "critical-level" utilitarianism. Assume that the utilitarian ethical decisionmaker makes the normative judgment that adding a single individual to the population at a "critical level" $(c^{crit}, \mathbf{a}^{crit}, R^{crit})$ is ethically neutral. She judges that, if $N(y) > N(x)$ and the added individuals all have this "critical-level" bundle, then the two outcomes are equally ethically good. Her view corresponds to the following SWF: $w(x) = \sum_{i=1}^{N(x)} v_i(x) - u(c^{crit}, \mathbf{a}^{crit}, R^{crit})$. Note that this approach is separable under certainty. Moreover, if the critical-level bundle is picked so as to have a higher well-being level than non-existence—that is, if $u(c^{crit}, \mathbf{a}^{crit}, R^{crit}) > u(c^{worth}, \mathbf{a}^{worth}, R^{worth})$—critical level utilitarianism satisfies the negative expansion principle *and* avoids the repugnant conclusion.

Actually, we can now see that the critical-level approach is an entire family of approaches, dependent on the choice of the critical bundle $(c^{crit}, \mathbf{a}^{crit}, R^{crit})$; and that "total utilitarianism" is the particular formula within this family that arises if we set the critical bundle at the level of a life worth living, $(c^{crit}, \mathbf{a}^{crit}, R^{crit}) = (c^{worth}, \mathbf{a}^{worth}, R^{worth})$. So we really have just two approaches, average vs "critical level" utilitarianism, and the further question for the latter approach of specifying the critical bundle.

Analogous formulas allow us to extend *prioritarianism* to the variable-population case. Average and critical-level *prioritarianism* are defined as follows:

$$\text{Average prioritarianism: } w(x) = \sum_{i=1}^{N(x)} \frac{g(v_i^*(x))}{N(x)}$$

$$\text{Critical level prioritarianism: } w(x) = \sum_{i=1}^{N(x)} g(v_i^*(x)) - g(u^*(c^{crit}, \mathbf{a}^{crit}, R^{crit}))$$

Total prioritarianism, in turn, is the version of critical-level prioritarianism that sets $(c^{crit}, \mathbf{a}^{crit}, R^{crit}) = (c^{worth}, \mathbf{a}^{worth}, R^{worth})$. The methodologies have features that parallel their utilitarian analogues. Average prioritarianism is non-separable even under certainty and violates the negative expansion principle. Total prioritarianism avoids these two difficulties, but has the "repugnant conclusion." Critical-level prioritarianism with a critical bundle better than non-existence is separable under certainty, satisfies the negative expansion principle, and avoids the repugnant conclusion.[43]

To a significant extent, then, the normative choices faced by prioritarianism and utilitarianism in the variable-population context are quite similar: the choice between the "average" or "critical level" approach and, if the latter, the identification of the critical-level bundle. However, it bears emphasis that, even in the fixed-population case, the prioritarian formula for the Atkinson SWF employs a well-being function, $v^*(.)$, which has already been "zeroed out" by identifying a special bundle, the zero-bundle $(c^{zero}, \mathbf{a}^{zero}, R^{zero})$. That is, $u^*(c, \mathbf{a}, R) = u(c, \mathbf{a}, R) - u(c^{zero}, \mathbf{a}^{zero}, R^{zero})$; $v_i^*(x) = u^*(c_i(x), \mathbf{a}_i(x), R_i(x))$. The critical-level prioritarian then generalizes the fixed-population Atkinson formula to the variable-population context by choosing a *second* special bundle, $(c^{crit}, \mathbf{a}^{crit}, R^{crit})$. By contrast, the formula for critical-level utilitarianism only requires a single special bundle, $(c^{crit}, \mathbf{a}^{crit}, R^{crit})$.

This contrast between the two critical-level approaches can be seen most clearly by expressing each in terms of a basal well-being function $u(.)$, unique up to an affine transformation.

$$\text{Critical level utilitarianism: } w(x) = \sum_{i=1}^{N(x)} u(c_i(x), \mathbf{a}_i(x), R_i(x)) - u(c^{crit}, \mathbf{a}^{crit}, R^{crit})$$

Critical level Atkinson prioritarianism: $w(x)$

$$= (1-\gamma)^{-1} \sum_{i=1}^{N(x)} \left(\left(u(c_i(x), \mathbf{a}_i(x), R_i(x)) - u(c^{zero}, \mathbf{a}^{zero}, R^{zero}) \right)^{1-\gamma} - \left(u(c^{crit}, \mathbf{a}^{crit}, R^{crit}) - u(c^{zero}, \mathbf{a}^{zero}, R^{zero}) \right)^{1-\gamma} \right)$$

To repeat: the *critical bundle* $(c^{crit}, \mathbf{a}^{crit}, R^{crit})$ is the level of ethical neutrality: ceteris paribus, changing an outcome by adding someone to the population with those attributes leaves the ethical value of the outcome unchanged. By contrast, *the zero bundle* $(c^{zero}, \mathbf{a}^{zero}, R^{zero})$ is the Atkinson point of "absolute priority," where the marginal moral impact of well-being and consumption go to infinity. Conceptually, these two thresholds are quite different from each other, and from a third threshold, the life just worth living $(c^{worth}, \mathbf{a}^{worth}, R^{worth})$. A life equally good as nonexistence need *not* be the point at which the marginal moral impact of well-being becomes infinite, nor the level at which adding someone to the population is ethically neutral. Moreover there are important pragmatic grounds for setting the zero bundle below the critical bundle, and indeed below the level of a life worth living. It is plausible (although, to be sure, a matter for normative discussion) that $u(c^{zero}, \mathbf{a}^{zero}, R^{zero}) < u(c^{worth}, \mathbf{a}^{worth}, R^{worth}) < u(c^{crit}, \mathbf{a}^{crit}, R^{crit})$.[44]

[43] Because outcomes in which anyone is assigned a negative well-being number are outside the domain of the Atkinson SWF, the Atkinson SWF can be said to "satisfy" these axioms only insofar as the axioms apply to an outcome set in which all well-being numbers are nonnegative. For example, if individuals with a life worse than nonexistence are assigned negative well-being numbers, an outcome in which some such individual is added to the population (as per the "negative expansion principle") is not within the Atkinson SWF's domain.

[44] By setting $u(c^{worth}, \mathbf{a}^{worth}, R^{worth}) < u(c^{crit}, \mathbf{a}^{crit}, R^{crit})$, we avoid the repugnant conclusion. Moreover, as already explained, because outcomes in which individuals have negative well-being numbers are outside the domain of the Atkinson SWF, there are pragmatic grounds for picking $(c^{zero}, \mathbf{a}^{zero}, R^{zero})$

We have thus far discussed variable population problems under conditions of certainty. The introduction of risk raises additional complications. In particular, the generalization of ex ante prioritarianism from the fixed- to the variable-population context is tricky. Ex ante prioritarianism requires assigning each person an expected well-being number; it is not obvious how to do this for individuals who have a non-zero probability of nonexistence. This topic cannot be pursued here.

By contrast, the extension of utilitarianism and ex post prioritarianism to the variable-population context is straightforward. Ex post critical-level prioritarianism uses the following formula (with analogous formulas for ex post average prioritarianism, ex post critical level utilitarianism, and ex post average utilitarianism):

$$W(a) = \sum_{s \in S} \pi_s H \left(\sum_{i=1}^{N(x_a^s)} g(v_i^*(x_a^s)) - g(u^*(c^{crit}, \mathbf{a}^{crit}, R^{crit})) \right)$$

Using an Atkinson SWF, and making explicit the normalization of well-being with the zero point, the formula becomes:

$$W(a) = \sum_{s \in S} \pi_s H \left((1-\gamma)^{-1} \sum_{i=1}^{N(x_a^s)} \left[\left(u\left(c_i(x_a^s), \mathbf{a}_i(x_a^s), R_i(x_a^s)\right) - u(c^{zero}, \mathbf{a}^{zero}, R^{zero})\right)^{1-\gamma} \right.\right.$$
$$\left.\left. - \left(u(c^{crit}, \mathbf{a}^{crit}, R^{crit}) - u(c^{zero}, \mathbf{a}^{zero}, R^{zero})\right)^{1-\gamma} \right] \right)$$

This is a plausible master formula, robust to both risk and to variation in population size; it simplifies to fixed-population ex post prioritarianism if $N(x) = N$ for all x, and to critical-level prioritarianism under certainty if the decisionmaker knows for certain what the state of nature s is. The formula has many attractions. By virtue of the concavity of $g(.)$, it satisfies the Pigou–Dalton principle, giving priority to worse-off individuals. It is time consistent and satisfies an axiom of stochastic dominance. It handles variable population in an attractive manner—satisfying the negative expansion principle and avoiding the repugnant conclusion, if the critical bundle is set above the level of a life worth living. With H itself concave, the formula is averse to catastrophic risks; alternatively, with H the identity function, the formula has nice separability properties. The formula provides a systematic basis for considering not only extinction risk, but the many other ways in which climate policies can affect population size.

That said, we must also keep clear in our minds the various normative judgments that are presupposed by this formula, or that are required to fully specify it. It is these choices that we hope climate scholarship will more explicitly engage and debate.

Footnote 44 continued
to be at or below the lowest possible well-being in all the outcomes under consideration. At the very least, if we wish to include in our evaluation outcomes in which individuals have lives no better than nonexistence, these pragmatic considerations argue for picking the zero bundle such that $u(c^{zero}, \mathbf{a}^{zero}, R^{zero}) \leq u(c^{worth}, \mathbf{a}^{worth}, R^{worth})$. Finally, note that setting $u(c^{zero}, \mathbf{a}^{zero}, R^{zero}) = u(c^{worth}, \mathbf{a}^{worth}, R^{worth})$ allows the Atkinson SWF to be well-defined at the level of a life worth living only for relatively low values of inequality aversion, $\gamma < 1$. However, it might be countered that setting $u(c^{zero}, \mathbf{a}^{zero}, R^{zero}) = u(c^{worth}, \mathbf{a}^{worth}, R^{worth})$ is intuitively "natural" (Adler 2012, ch. 3).

7 Conclusion

Prioritarianism is an ethical view that gives greater weight to the changes in well-being affecting individuals (or generations) at lower well-being levels. Axiomatically, this view differs from utilitarianism in conforming to the Pigou–Dalton principle, which has much intuitive force. Technically, one obtains a prioritarian SWF by summing a concave transformation of individual well-being numbers. Prioritarianism is thus morally intuitive, well-grounded axiomatically and fairly parsimonious in terms of economic modeling.

In this Article, we have described the range of normative questions/puzzles that would typically arise in specifying a prioritarian SWF. In particular, we have discussed how to measure and scale the individual well-being functions, and how to apply a prioritarian SWF under risk and with endogenous variable populations. These issues are not fully settled in the social choice literature, and we have tried to systematically represent the various difficulties and trade-offs.

Our main objective in this Article was to prepare the ground for a systematic exploration of the implications of prioritarianism for climate change policies. In the course of our discussion, we have nevertheless discussed some of these possible implications, with respect to the social discount rate, the social cost of carbon, optimal mitigation, and the dismal theorem. We hope that this discussion will motivate scholars working on climate change to give prioritarianism serious consideration in the future.

Acknowledgments This article grows out of presentations by the authors at conferences on climate change held at Duke University, Goteborg University, and Princeton University; at the conference of the European Association of Environmental and Resource Economists (EAERE); and at the Snowmass modelling workshop. Thanks to participants in those events for their comments, and to two anonymous referees. Nicolas Treich acknowledges funding from the chair "Finance Durable et Investissement Responsable" (FDIR).

References

Adler MD (2009) Future generations: a prioritarian view. George Wash Law Rev 77:1478–1519

Adler MD (2012) Well-being and fair distribution: beyond cost-benefit analysis. Oxford University Press, New York

Adler MD (2016) Extended preferences. In: Adler MD, Fleurbaey M (eds) Oxford handbook of well-being and public policy. Oxford University Press, New York (forthcoming)

Adler MD, Hammitt JK, Treich N (2014) The social value of mortality risk reduction: VSL versus the social welfare function approach. J Health Econ 35:82–93

Adler, MD, Treich N (2014) Consumption, risk and prioritarianism. SSRN working paper no. 2508440

Anthoff D, Hepburn C, Tol RSJ (2009) Equity weighting and the marginal damage costs of climate change. Ecol Econ 68:836–849

Arneson RJ (2007) Desert and equality. In: Holtug N, Lippert-Rasmussen K (eds) Egalitarianism: new essays on the nature and value of equality. Clarendon Press, Oxford, pp 262–293

Arrhenius G (2016) Population ethics. Oxford University Press, Oxford (forthcoming)

Arrow KJ et al (2014) Should governments use a declining discount rate in project analysis? Rev Environ Econ Policy 8:145–163

Bergson A (1948) Socialist economics. In: Ellis HS (ed) A survey of contemporary economics, vol 1. Richard D. Irwin, Homewood, pp 412–448

Bergson A (1954) On the concept of social welfare. Quart J Econ 68:233–252

Blackorby C, Bossert W, Donaldson D (2000) The value of limited altruism. J Econ Theory 95:37–70

Blackorby C, Bossert W, Donaldson D (2005) Population issues in social choice theory, welfare economics, and ethics. Cambridge University Press, Cambridge

Boadway RW, Bruce N (1984) Welfare economics. Blackwell, Oxford

Bommier A, Zuber S (2008) Can preferences for catastrophe avoidance reconcile social discounting with intergenerational equity? Soc Choice Welfare 31:415–434

Bossert W, Weymark JA (2004) Utility in social choice. In: Barberà S, Hammond PJ, Seidl C (eds) Handbook of utility theory, vol. 2 (extensions). Kluwer, Boston, pp 1099–1177
Botzen WJW, van den Bergh JCJM (2014) Specifications of social welfare in economic studies of climate policy: overview of criteria and related policy insights. Environ Resour Econ 58:1–33
Brown C (2005) Priority or sufficiency... or both? Econ Philos 21:199–220
Broome J (2004) Weighing lives. Oxford University Press, Oxford
Broome J (2008) The ethics of climate change. Sci Am 2008:97–102
Carlsson F, Daruvala D, Johansson-Stenman O (2005) Are people inequality-averse or just risk-averse? Economica 72:375–396
Dasgupta P (2008) Discounting climate change. J Risk Uncertain 37:141–169
Dasgupta P (2012) Time and the generations. In: Hahn RW, Ulph A (eds) Climate change and common sense: essays in honour of Tom Schelling. Oxford University Press, Oxford, pp 101–130
Dietz S, Asheim G (2012) Climate policy under sustainable discounted utilitarianism. J Environ Econ Manag 63:321–335
Drèze J, Stern N (1987) The theory of cost-benefit analysis. In: Auerbach AJ, Feldstein M (eds) Handbook of public economics, vol 2. North-Holland, Amsterdam, pp 909–989
Farber DA (2003) From here to eternity: environmental law and future generations. Univ Illinois Law Rev 2003:289–336
Fleurbaey M (2010) Assessing risky social situations. J Polit Econ 118:649–680
Fleurbaey M (2016) Equivalent income. In: Adler MD, Fleurbaey M (eds) Oxford handbook of well-being and public policy. Oxford University Press, New York (forthcoming)
Fleurbaey M, Bovens L (2012) Evaluating life or death prospects. Econ Philos 28:217–249
Gaertner W, Schokkaert E (2012) Empirical social choice: questionnaire-experimental studies on distributive justice. Cambridge University Press, Cambridge
Gollier C (2001) The economics of risk and time. MIT Press, Cambridge, MA
Gollier C (2013) Pricing the planet's future: the economics of discounting in an uncertain world. Princeton University Press, Princeton
Gollier C, Weitzman ML (2010) How should the distant future be discounted when discount rates are uncertain? Econ Lett 107:350–353
Grant S, Kajii A, Polak B, Safra Z (2010) Generalized utilitarianism and Harsanyi's impartial observer theorem. Econometrica 78:1939–1971
Greenstone M, Kopits E, Wolverton A (2013) Developing a social cost of carbon for US regulatory analysis: a methodology and interpretation. Rev Environ Econ Policy 7:23–46
Groom B et al (2005) Declining discount rates: the long and the short of it. Environ Resour Econ 32:445–493
Hammond P (1983) Ex-post optimality as a dynamically consistent objective for collective choice under uncertainty. In: Pattanaik PK, Salles M (eds) Social choice and welfare. North-Holland, Amsterdam, pp 175–205
Harsanyi JC (1977) Rational behavior and bargaining equilibrium in games and social situations. Cambridge University Press, Cambridge
Holtug N (2010) Persons, interests, and justice. Oxford University Press, Oxford
Holtug N (2015) Theories of value aggregation: utilitarianism, egalitarianism, prioritarianism. In: Hirose I, Olson J (eds) The Oxford handbook of value theory. Oxford University Press, Oxford, pp 267–284
Intergovernmental Panel on Climate Change (IPCC) Working Group III (2014) Social, economic and ethical concepts and methods. In: Climate change 2014: mitigation of climate change
Kagan S (1989) The limits of morality. Clarendon Press, Oxford
Kaplow L (2008) The theory of taxation and public economics. Princeton University Press, Princeton
Kaplow L (2010) Concavity of utility, concavity of welfare, and redistribution of income. Int Tax Public Finance 17:25–42
Kaplow L, Weisbach D (2011) Discount rates, social judgments, individuals' risk preferences, and uncertainty. J Risk Uncertain 42:125–143
Keeney RL (1980) Equity and public risk. Oper Res 28:527–534
Koopmans TC (1960) Stationary ordinal utility and impatience. Econometrica 28:287–309
Kreps DM (1988) Notes on the theory of choice. Westview Press, Boulder
Llavador H, Roemer JE, Silvestre J (2011) A dynamic analysis of human welfare in a warming planet. J Public Econ 95:1607–1620
McKerlie D (2007) Egalitarianism and the difference between interpersonal and intrapersonal judgments. In: Holtug N, Lippert-Rasmussen K (eds) Egalitarianism: new essays on the nature and value of equality. Clarendon Press, Oxford, pp 157–173
Millner A (2013) On welfare frameworks and catastrophic climate risks. J Environ Econ Manag 65:310–325
Millner A, Dietz S, Heal G (2013) Scientific ambiguity and climate policy. Environ Resour Econ 55:21–46

Neumayer E (2013) Weak versus strong sustainability: exploring the limits of two opposing paradigms. Edward Elgar, Cheltenham
Nordhaus WD (2007) A review of the Stern review on the economics of climate change. J Econ Lit 45:686–702
Nordhaus WD (2008) A question of balance: weighing the options on global warming policies. Yale University Press, New Haven
Nordhaus WD (2010) Economic aspects of global warming in a post-Copenhagen environment. Proc Natl Acad Sci 107:11721–11726
Parfit D (1987) Reasons and persons. Revised and corrected paper edition. Oxford University Press, Oxford
Parfit D (1991) Equality or priority. Lindley lecture. University of Kansas. Reprinted in M. Clayton and A. Williams (eds) The Ideal of Equality (Houndmills: Palgrave, 2000), pp 81–125
Parfit D (2012) Another defence of the priority view. Utilitas 24:399–440
Pindyck RS (2013) Climate change policy: what do the models tell us? J Econ Lit 51:860–872
Porter T (2012) In defence of the priority view. Utilitas 24:197–206
Ramsey FP (1928) A mathematical theory of saving. Econ J 38:543–559
Roberts MA (2015) Population axiology. In: Hirose I, Olson J (eds) The Oxford handbook of value theory. Oxford University Press, Oxford, pp 399–423
Roemer JE (2010) The ethics of intertemporal distribution in a warming planet. Environ Resour Econ 48:363–390
Samuelson PA (1947) Foundations of economic analysis. Harvard University Press, Cambridge
Scheffler S (1982) The rejection of consequentialism. Clarendon Press, Oxford
Schelling TC (1995) Intergenerational discounting. Energy Policy 23:395–401
Sidgwick H (1907) The methods of ethics, 7th edn. Macmillan, London. Reprinted 1981 (Indianapolis: Hackett)
Stern N (2007) The economics of climate change: the Stern review. Cambridge University Press, Cambridge
Sterner T, Persson UM (2008) An even Sterner review: introducing relative prices into the discounting debate. Rev Environ Econ Policy 2:61–76
Sumner LW (1996) Welfare, happiness and ethics. Clarendon Press, Oxford
Tol RSJ (2011) The social cost of carbon. Annu Rev Resour Econ 3:419–443
Traeger CP (2013) Discounting under uncertainty: disentangling the Weitzman and the Gollier effect. J Environ Econ Manag 66:573–582
Tungodden B (2003) The value of equality. Econ Philos 19:1–44
van den Bergh JCJM, Botzen WJW (2014) A lower bound to the social cost of CO_2 emissions. Nat Clim Change 4:253–258
von Weizsacker CC (1965) Existence of optimal programs of accumulation for an infinite time horizon. Rev Econ Stud 32:85–104
Weitzman ML (1998) Why the far-distant future should be discounted at its lowest possible rate. J Environ Econ Manag 36:201–208
Weitzman ML (2001) Gamma discounting. Am Econ Rev 91:260–271
Weitzman ML (2007) A review of the Stern review on the economics of climate change. J Econ Lit 45:703–724
Weitzman ML (2009) On modelling and interpreting the economics of catastrophic climate change. Rev Econ Stat 91:1–19
Weitzman M (2011) Fat-tailed uncertainty in the economics of catastrophic climate change. Rev Environ Econ Policy 5:275–292
Weymark JA (2016) Social welfare functions. In: Adler MD, Fleurbaey M (eds) Oxford handbook of well-being and public policy. Oxford University Press, New York (forthcoming)
Williams A (2012) The priority view bites the dust? Utilitas 24:315–331