# Future Generations: A Prioritarian View

Matthew D. Adler*

How should we take account of the interests of future generations? This question has great practical relevance. For example, it is front and center in arguments about global warming policy. Unfortunately, the question is doubly difficult—*doubly*, because it not merely implicates generic disputes about the structure of moral obligations (disputes between consequentialists and nonconsequentialists, welfarists and nonwelfarists, and so forth), but because the appropriate treatment of future generations implicates distinct problems that are not resolved merely by adopting one or another generic framework.

This Article addresses the appropriate treatment of future generations within the framework of *prioritarianism*. In prior articles and a forthcoming book, I argue that the "social welfare function" ("SWF") approach to policy choice provides a systematic, implementable, and theoretically well-grounded method for rendering policy sensitive not only to efficiency, but also to equity—to the fair distribution of well-being.[1] In particular, I defend a "prioritarian" SWF. A prioritarian SWF ranks outcomes using the formula $\sum_{i=1}^{N} g(u_i(x))$—where $x$ is an outcome; $u_i(x)$ an individual utility number representing the well-being of individual $i$ in outcome $x$; and $g(.)$ a strictly increasing and strictly concave function.[2] Because $g(.)$ is strictly increasing and strictly concave, changes to the well-being of worse-off individuals have greater social weight than changes to the well-being of better-off individuals. The prioritarian SWF is therefore sensitive to eq-

* Leon Meltzer Professor, University of Pennsylvania Law School. Many thanks for their helpful comments on this Article, or conversations about problems of future generations, to Seth Baum, Bill Buzbee, Dan Farber, Robert Hockett, Doug Kysar, Eric Posner, Arden Rowell, Chris Sanchirico, David Weisbach, and the participants in *The George Washington Law Review* symposium at which the Article was presented in draft form.

1 *See* MATTHEW D. ADLER, WELL-BEING AND EQUITY: A FRAMEWORK FOR POLICY ANALYSIS (forthcoming 2010) [hereinafter, ADLER, WELL-BEING AND EQUITY]; Matthew D. Adler & Chris William Sanchirico, *Inequality and Uncertainty: Theory and Legal Applications*, 155 U. PA. L. REV. 279 (2006); Matthew D. Adler, *Risk Equity: A New Proposal*, 32 HARV. ENVTL. L. REV. 1 (2008); Matthew D. Adler, *Well-Being, Inequality and Time: The Time-Slice Problem and Its Policy Implications* (Univ. of Pa. Law Sch. Inst. for Law & Econ., Working Paper No. 07-17, 2007), *available at* http://ssrn.com/abstract=1006871.

2 This is the appropriate formula for what I call the "core case," where the very same $N$ individuals exist in all outcomes. Variations in the formula for other cases are considered below. *See infra* Part III.

**September 2009 Vol. 77 No. 5/6**

uity.[3] This framework has firm intellectual roots, drawing both from moral philosophy (where the distinction between prioritarian and non-prioritarian approaches to equity has recently been much discussed), and from welfare economics (where the idea of an SWF originates).

My claim, in this Article, is that the appropriate treatment of future generations, within the framework of prioritarianism, is straightforward in the "core case" where the intertemporal population is fixed and finite. In other words, the same $N$ individuals exist in all of the possible outcomes[4] of the policy choice at hand, with $N < \infty$. Things become trickier if we relax the assumption of a fixed, finite population. One departure from the "core case" involves *non-identity* problems—where there are individuals who exist in some, but not all, of the possible outcomes. In other words, there is at least one pair of possible outcomes $x$ and $y$, such that individual $i$ exists in $x$ but not in $y$. A related but logically distinct problem involves *variation in the size of populations*. In other words, there is at least one pair of possible outcomes $x$ and $y$, such that outcome $x$ has $N$ individuals, outcome $y$ has $M$ individuals, and $N \neq M$. Finally, *infinity* problems arise where some or all of the possible outcomes contain an infinite number of individuals. This can happen, in particular, if each generation contains a finite number of individuals but time continues indefinitely.

Part I of this Article summarizes the prioritarian approach. Part II discusses the core case. Here, I shall argue that the interests of future generations should be given the very same weight as the interests of the present generation. Arguments to the effect that the SWF should incorporate a utility discount factor, reflecting a pure rate of time preference, are misconceived—at least in the core case.

Part III discusses departures from the core case: non-identity problems, variation in the size of populations, and infinite populations. As we shall see, arguments for departing from neutrality between present and future generations become stronger in such cases. In particular, there is a plausible argument that an individual who does not exist in some outcomes should have her interests wholly ignored for purposes of comparing such outcomes to outcomes in which she does exist. As for infinity cases, we shall see that a strong kind of

---

[3] Formally, a SWF is sensitive to equity if it satisfies the "Pigou-Dalton" principle. Prioritarian SWFs are a subset of the SWFs that satisfy this principle. *See* sources cited *supra* note 1.

[4] By "possible outcomes," I mean all the outcomes in the outcome set, $O$, which is relevant to the choice at hand. *See infra* text accompanying notes 13–15.

temporal neutrality is logically inconsistent with the principle of Pareto superiority.

Part III tentatively concludes that neutrality *should* be maintained in non-identity and population-size-variation cases. It suggests that these cases might be handled via a simple generalization of the prioritarian formula. Outcome $x$ is ranked as being at least as good as outcome $y$ iff[5] $\sum_{i=1}^{N(x)} g(u_i(x)) \geq \sum_{i=1}^{N(y)} g(u_i(y))$ —where $N(x)$ is the number of individuals who exist in $x$, $N(y)$ in outcome $y$. The effect of this formula is to take full account of the interests of potential nonexistents in comparing outcomes where they do and do not exist.[6] As for infinite-population cases, Part III suggests that some measure of temporal neutrality might be preserved in such cases via a prioritarian variation on the "overtaking" criterion.

These conclusions, however, are much less firm and definitive than the argument for neutrality in the core case. The main aim of Part III is simply to acquaint the reader with the various kinds of problems for the SWF framework that arise once we relax the assumptions of a fixed, finite population. Given the intrinsic difficulty of these problems, and space limitations, I can hardly do more than that.

## *I. Prioritarian SWFs*

The possibility of interpersonal welfare comparisons has long been, and remains, a controversial issue among welfare economists. Many economists still reject such comparisons and espouse policy-analytic frameworks that do not require them—for example, the criterion of Kaldor-Hicks efficiency. However, a substantial number of economists take the position that well-being is indeed comparable across persons. These economists tend to favor the use of SWFs to evaluate policy.

There is a large theoretical literature concerning SWFs,[7] and also more "applied" literatures, where economists deploy SWFs to evalu-

---

5 "Iff" means "if and only if."

6 If individual $i$ exists in outcome $x$ but not $y$, and has utility $u_i(x)$ in $x$, the effect of the above formula is to assign $i$ a utility level of zero in $y$, and to take full account of the difference between $u_i(x)$ and zero in ranking the two outcomes, as much as if $i$ existed in outcome $y$.

7 For reviews of this literature, see Walter Bossert & John A. Weymark, *Utility in Social Choice*, *in* 2 HANDBOOK OF UTILITY THEORY 1099 (Salvador Barberà et al. eds., 2004); and Claude d'Aspremont & Louis Gevers, *Social Welfare Functionals and Interpersonal Comparability*, *in* 1 HANDBOOK OF SOCIAL CHOICE AND WELFARE 459 (Kenneth J. Arrow et al. eds., 2002).

ate particular kinds of policies. For example, the whole scholarly field of "optimal tax policy" revolves around the use of SWFs.[8] SWFs are also increasingly used in environmental economics.[9] Much economic analysis of climate change employs the SWF framework—including both the recent Stern report[10] and competing work (such as well-known work by William Nordhaus[11]). Nordhaus and Stern disagree, vigorously, about the choice of discount rate,[12] but concur in looking at the problem of global warming through the lens of an SWF.

In prior articles and a forthcoming book, I draw upon both the various bodies of economic scholarship just referenced and contemporary scholarship in moral philosophy, and argue for a prioritarian SWF: one that has the form $\sum_{i=1}^{N} g(u_i(x))$.[13] Here, I will very briefly summarize the main elements of this prioritarian framework. The reader is referred to the work just cited for a much fuller explication and defense.[14]

The prioritarian framework is, concededly, *consequentialist* and *welfarist*. A policymaker faces some set of possible choices $A = \{a, b, c, \ldots\}$. The prioritarian framework tells the decisionmaker to think about those choices by, first, thinking about the possible outcomes of the choices and ranking those outcomes. Formally, a given choice situation is matched with an outcome set $O = \{x, y, z, \ldots\}$, where each element of $O$ is a description of a possible way the world might turn out. The function of the prioritarian SWF is to rank the outcomes in $O$ and thereby—ultimately—to determine the ranking of the choices in $A$.

Because the prioritarian framework is meant to function as a

---

8 For reviews, see MATTI TUOMALA, OPTIMAL INCOME TAX AND REDISTRIBUTION (1990); Christopher Heady, *Optimal Taxation as a Guide to Tax Policy*, *in* THE ECONOMICS OF TAX POLICY 23 (Michael P. Devereux ed., 1996); Nicholas Stern, *The Theory of Optimal Commodity and Income Taxation: An Introduction*, *in* THE THEORY OF TAXATION FOR DEVELOPING COUNTRIES 22 (David Newbery & Nicholas Stern eds., 1987).

9 *See* Ian W.H. Parry et al., *The Incidence of Pollution Control Policies* 26–28 (Nat'l Bureau of Econ. Research, Working Paper No. 11438, 2005).

10 NICHOLAS STERN, THE ECONOMICS OF CLIMATE CHANGE (2007).

11 WILLIAM NORDHAUS, A QUESTION OF BALANCE (2008).

12 *See* NORDHAUS, *supra* note 11, at 165–91; David Weisbach & Cass R. Sunstein, *Climate Change and Discounting the Future: A Guide for the Perplexed* (Reg-Markets Ctr., Working Paper No. 08-19, 2008), *available at* http://ssrn.com/abstract=1223448.

13 *See* sources cited *supra* note 1. The articles cited there are agnostic as between prioritarianism and other types of equity-regarding SWFs, but my forthcoming book defends prioritarianism.

14 *See id.*

workable tool for policy choice—a tool that should be useable by actual human decisionmakers, with "bounded" cognitive abilities—outcomes are *simplified* descriptions of ways the world might turn out, not complete "possible worlds." An outcome describes what might occur, with respect to some of the determinants of human well-being. For example, each outcome in $O$ might be a different possible realization of annual consumption amounts for each member of the population. Or, each might consist of a different possible realization of annual health states for each member of the population. Or, each outcome might describe how each individual fares with respect to both health and consumption, or how she fares with respect to both consumption and various public goods.

Questions about the appropriate construction and simplification of outcome sets have great practical relevance, and I do address them at length elsewhere.[15] But they are too complicated to discuss in brief compass, and are orthogonal to the issues of this Article—and so I ignore them here. For purposes of the analysis that follows, all I need to assume is that each choice situation is paired with an outcome set $O$, containing descriptions of possible ways the world might be. Each such description—each element of $O$—may be more or less simplified, but does include information about which individuals exist.

The prioritarian framework is *consequentialist* because it derives a ranking of policy choices from a ranking of outcomes. It is *welfarist* because it makes the ranking of outcomes solely a function of human well-being. Some readers will find these features of the framework attractive. Others will not—whether because they reject consequentialism, or because they accept consequentialism but reject the premise that the ranking of outcomes should be solely a function of human well-being. Even some of these readers, however, might come to see the prioritarian framework as a useful approximation, or as one component in a broader moral theory.[16]

The prioritarian framework, like other approaches in the SWF family, presupposes the possibility of interpersonal welfare comparisons. It represents such comparisons via the device of a utility func-

---

15 *See* ADLER, WELL-BEING AND EQUITY, *supra* note 1.

16 Theorists who are concerned about equalizing *opportunity for well-being*, rather than well-being itself, may see prioritarianism as a rough approximation of the correct moral theory. Hybrid theorists who believe that morality requires the maximization of good consequences, within deontological constraints, may see prioritarianism as one component of a broader moral theory—as may theorists who believe that morality requires attention both to human well-being, and to other considerations, e.g., animal well-being or intrinsic environmental values. *See id.*

tion $u(.)$, which helps to determine the ranking of outcomes. The utility function maps a given outcome, $x$, onto a vector of individual utilities.

In the core case where the very same $N$ individuals exist in every outcome in $O$, $u(.)$ maps $x$ onto the vector $(u_1(x), u_2(x), \ldots, u_N(x))$. $u_1(x)$ is the utility of individual 1 in outcome $x$, $u_2(x)$ the utility of individual 2 in outcome $x$, and so forth. These utility numbers allow for interpersonal comparisons of well-being levels; for interpersonal comparisons of well-being differences; and for comparisons of each individual's well-being to a "neutral" level of zero well-being. In other words, $u_i(x) > u_j(y)$ iff individual $i$ is better off in outcome $x$ than individual $j$ is in outcome $y$. Further, $|u_i(x) - u_i(y)| > |u_j(w) - u_j(z)|$ iff the difference between individual $i$'s well-being in $x$ and $y$ is greater than the difference between individual $j$'s well-being in $w$ and $z$. Finally, $u_i(x) > 0$ iff $i$ is better off in $x$ than the neutral level.

But how *is* it possible to compare welfare levels and differences across persons; to compare welfare to a zero level; and to construct a numerical utility function $u(.)$ that faithfully represents these comparisons? Here, I draw upon John Harsanyi's notion of *extended preferences*.[17] An extended preference is a preference to be one or another person. Formally, extended preferences are preferences regarding *life histories*. A life history takes the form $(x; i)$, which means being individual $i$ in outcome $x$. Assume once more the core case of $N$ individuals who exist in all outcomes in the outcome set $O = \{x, y, \ldots\}$, with $K$ outcomes in total. Then there are $N \times K$ life histories. For each of the $N$ individuals, we can ask about her extended preferences over the $N \times K$ life histories—not her actual extended preferences but rather, I suggest, her fully informed, fully rational, self-interested extended preferences. Further, we can ask about her fully informed, fully rational, self-interested extended preferences regarding different lotteries over the $N \times K$ life histories. Finally, we can ask about her fully informed, fully rational, self-interested extended preferences as between various life histories and the prospect of non-existence. And we can represent this extended preference structure, for the given individual, via a utility function—at least if the individual's fully informed, fully rational, self-interested extended preferences are complete.

---

17 *See* JOHN C. HARSANYI, RATIONAL BEHAVIOR AND BARGAINING EQUILIBRIUM IN GAMES AND SOCIAL SITUATIONS 48–83 (1977); John A. Weymark, *A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism*, *in* INTERPERSONAL COMPARISONS OF WELL-BEING 255, 289–97 (Jon Elster & John E. Roemer eds., 1991).

Harsanyi makes the further assumption that each of the $N$ individuals will have the very same extended preferences over life histories and lotteries, and that these extended preferences will indeed be complete. Both assumptions are problematic, in my view, and in my explication of the prioritarian framework I discuss how they might be relaxed.[18] For purposes of this Article, however, the possibility of incomplete or divergent extended preferences is a complication that can be ignored. Assume, for simplicity, that the $N$ individuals, if fully informed, fully rational, and self-interested, would have the same, complete ranking of life histories and lotteries over life histories, and would converge in how they compare each life history to nonexistence. This single, complete extended-preference structure can be represented by a utility function $u(x; i)$, which will be unique, up to a ratio transformation.[19] Because $u(x; i)$ is unique up to a ratio transformation, it contains sufficient information to make interpersonal comparisons of both well-being levels and differences, and to allow us to say whether a given life history is above or below the zero level.

In short, we need an interpersonally comparable utility function for our SWF, $u(.)$, which will map each outcome onto a vector $(u_1(x), u_2(x), \ldots, u_N(x))$; and we can produce that function by defining $u_i(x) = u(x; i)$, the utility of having the life history of person $i$ in outcome $x$.

Generically, a SWF takes the form $w(x) = w(u_1(x), u_2(x), \ldots, u_N(x))$. It attaches a social value to each outcome $w(x)$, as a function of the individual utilities in that outcome $(u_1(x), u_2(x), \ldots, u_N(x))$. This SWF value, $w(x)$, is meant to *represent* the ranking of the outcomes. $w(x) \geq w(y)$ iff $x$ is at least as good[20] an outcome as $y$.

Prioritarianism uses a particular functional form for $w(x)$, namely $w(x) = \sum_{i=1}^{N} g(u_i(x))$. Why adopt this particular formula for ranking outcomes?

Philosophers who have explicated prioritarianism agree that its distinctive feature is this: while utilitarianism simply adds up individual utilities, prioritarianism gives extra moral weight to well-being changes affecting individuals at lower well-being levels. Those who

---

18 *See* ADLER, WELL-BEING AND EQUITY, *supra* note 1; *see also* MATTHEW D. ADLER & ERIC A. POSNER, NEW FOUNDATIONS OF COST-BENEFIT ANALYSIS 47–50 (2006).

19 This means that $v(.)$ produces the very same ranking of life histories, lotteries over life histories, and comparisons of each life history to nonexistence as $u(.)$ iff $v(.) = ku(.)$, where $k$ is a positive constant.

20 "At least as good as" means either better than or equally good as.

have discussed how to represent prioritarianism through the SWF formalism also agree that the form $\sum_{i=1}^{N} g(u_i(x))$ is the way to do so.[21] The strict increasingness and concavity of $g(.)$ has precisely the implication that the change in the value of the SWF, produced by a given change in an individual's utility, decreases as the individual's utility level increases. Consider that, if individual $i$ is at utility level $u_i(x)$ in outcome $x$, the change in the value of the SWF[22] produced by increasing $i$'s utility by a positive amount $\Delta u$ is $g(u_i(x) + \Delta u) - g(u_i(x))$. Consider now the change in the value of the SWF produced by increasing individual $j$'s utility by the same $\Delta u$ amount, where $j$ is worse off than $i$: $u_j(x) < u_i(x)$. Because $g$ is strictly increasing and strictly concave, this change will be greater in $j$'s case than in $i$'s.[23]

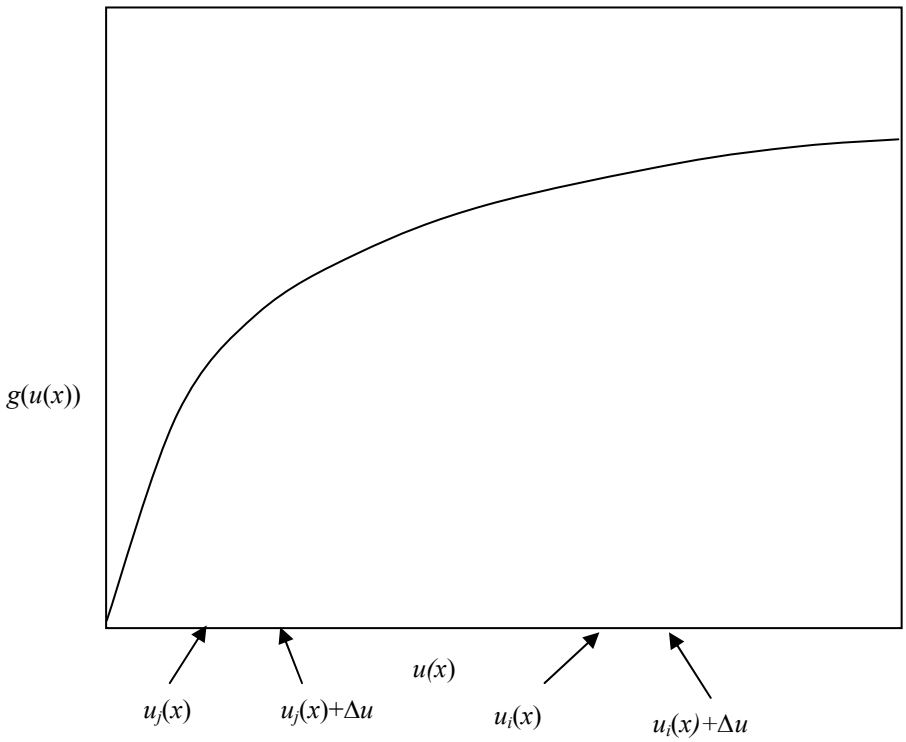
---

21 *See* Karsten Klint Jensen, *What Is the Difference between (Moderate) Egalitarianism and Prioritarianism?*, 19 ECON. & PHIL. 89, 99 (2003); John Broome, *Equality Versus Priority: A Useful Distinction*, *in* FAIRNESS AND GOODNESS IN HEALTH (Daniel Wikler et al. eds., forthcoming) (manuscript at 2, on file with author); Nils Holtug, *Prioritarianism*, *in* EGALITARIANISM: NEW ESSAYS ON THE NATURE AND VALUE OF EQUALITY 125, 133 (Nils Holtug & Kasper Lippert-Rasmussen eds., 2007); Wlodek Rabinowicz, *Prioritarianism and Uncertainty: On the Interpersonal Addition Theorem and the Priority View*, *in* EXPLORING PRACTICAL PHILOSOPHY: FROM ACTION TO VALUES 139, 148 (Dan Egonsson et al. eds., 2001); Campbell Brown, Matters of Priority 99–105 (Mar. 2005) (unpublished Ph.D. thesis, The Australian National University) (on file with author).

22 "The value of the SWF" means $w(x) = \sum_{i=1}^{N} g(u_i(x))$.

23 *See* Brown, *supra* note 21 at 117–18.

**Chart 1. Transforming Utility with a Strictly Increasing and Concave *g*-Function.**




The question remains, however: why *do* well-being changes affecting individuals at lower well-being levels have greater moral weight? Here I offer the following picture, building on work by Thomas Nagel.[24] Given two outcomes $x$ and $y$, the moral ranking of $x$ and $y$ should be a function of individuals' claims in favor of $x$ or $y$. "Claims," in this picture, are relations between an individual and a pair of outcomes. As between outcomes $x$ and $y$, a given individual has a claim in favor of $x$ over $y$, a claim in favor of $y$ over $x$, or no claim either way.

What leads to prioritarianism are the following substantive propositions concerning claims. If individual $i$ is equally well off in $x$ and $y$, then she has no claim either way. If $i$ is better off in $x$ than $y$, then she has a claim in favor of $x$ over $y$. Further, in the case where an individual is better off in one outcome than another, the *strength* of

24 *See* THOMAS NAGEL, EQUALITY AND PARTIALITY 63–74 (1991) (presenting a conception of equality that sees each individual as having a distinct claim to well-being, and that gives priority to worse-off individuals); THOMAS NAGEL, MORTAL QUESTIONS 106–27 (1979) (same).

her claim for the first outcome depends both on her well-being difference between the two outcomes, and on her well-being levels in the two outcomes. The strength of the claim varies directly with the well-being difference and inversely with the levels.[25] In particular, consider the case where $i$ is better off in $x$ than $y$, and $j$ is better off in $y$ than $x$; the well-being differences are the same;[26] but $j$ is worse off in both outcomes than $i$ is in both outcomes.[27] In this case (at least if $j$ is not responsible for being worse off than $i$ in both outcomes), it is only fair that we count $j$ as having a stronger claim between the outcomes than $i$.[28]

This approach to ranking pairs of outcomes—as a function of individuals' claims in favor of one outcome or the other, with worse-off individuals having stronger claims—leads us to the prioritarian SWF. I should also note that it helps resolve a different question: namely, whether the ranking of outcomes should be a function of individuals' lifetime well-being in the outcomes, or instead a function of individuals' well-being during "sublifetime" periods (e.g., annual well-being). I adopt the lifetime approach. $u_i(x)$ is a utility number representing $i$'s *lifetime* well-being in outcome $x$. The SWF formula that I defend, $\sum_{i=1}^{N} g(u_i(x))$, means the sum of a strictly increasing and strictly concave transformation of individuals' lifetime utilities. This approach is justified by the fact that human personhood (normally) endures over an entire lifetime. Each human being (normally) remains one and the same person from the time she is born until the end of her human life. The deep rationale for ranking pairs of outcomes by considering individuals' claims for or against the outcomes is that each individual is *distinct* from every other—a different particular person, with a distinct moral "vote" to determine how the world should be shaped. And, because personhood endures over a human lifetime, the "currency" for individuals' claims should be lifetime rather than sublifetime well-being.

The prioritarian SWF, $\sum_{i=1}^{N} g(u_i(x))$, satisfies a number of axioms.

---

25 Of course, the individual is at different levels in the two outcomes, so to be more precise, we should say that the strength of the claim is a decreasing function of the average well-being level.

26 That is, $u_i(x) - u_i(y) = u_j(y) - u_j(x)$.

27 That is, $u_i(x) > u_i(y) > u_j(y) > u_j(x)$.

28 Louis Kaplow and Steven Shavell use the term "fairness" to mean non-welfarism. *See* LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE 38–45 (2002). However, if one views fairness (as I do) as a matter of the equitable satisfaction of individuals' claims to well-being, there is no inconsistency between fairness and welfarism.

All of these axioms can be given a strong substantive justification in the theory of prioritarianism (at least in the claim-based version as I have just summarized it).

> *Pareto Indifference*: If each individual is just as well off in outcome *x* as she is in *y*, then the two outcomes should be ranked by the SWF as equally good.
>
> *Anonymity*: If the utility levels in *x* are just a rearrangement ("permutation") of the utility levels in *y*, *x* and *y* should be ranked as equally good.[29]
>
> *Pareto Superiority*: If each individual is at least as well off in *x* as in *y*, and one or more individuals are strictly better off in *x*, then *x* should be ranked as better than *y*.
>
> *Pigou-Dalton*: If *i* is better off than *j* in some outcome, and we transfer a fixed amount of utility from *i* to *j*, leaving everyone else unaffected, and leaving *j* still worse off than *i*, the new outcome should be ranked as better than the original one.
>
> *Separability-Across-Individuals*: If some individuals are just as well off in *x* and *y*, the ranking of *x* and *y* does not depend on what particular level of well-being each individual attains in the two outcomes.

Even if we accept the basic argument for a prioritarian SWF, more room for discussion remains. After all, there are many different kinds of strictly increasing and strictly concave *g*-functions. I believe that the "Atkinsonian" class of SWFs—a subfamily within the broader prioritarian family—provides the most attractive basis for ranking outcomes. The Atkinsonian SWFs have the form $\frac{1}{1-\gamma}\sum_{i=1}^{N} u_i(x)^{1-\gamma}$. They are parameterized by a single inequality-aversion parameter $\gamma$, which ranges from 0 to infinity. The larger the value of $\gamma$, the greater the weight given to welfare changes affecting worse-off individuals. With $\gamma = 0$, the SWF is no longer prioritarian and becomes utilitarian. As $\gamma$ approaches infinity, the SWF approaches the "leximin" SWF.[30]

---

29 In the case of infinite populations, we need to distinguish between different variants of the anonymity axiom. *See infra* note 84 and accompanying text. But in the finite case, anonymity is straightforward: utility vectors with the same, finite number of entries that are permutations of each other must be ranked as equally good.

30 Leximin says to rank two outcomes equally if the utility vectors corresponding to the two are permutations of each other. Otherwise, *x* is better than *y* if the worst-off individual in *x* is better off than the worst-off individual in *y*; if their well-being levels are equal, then *x* is better

Atkinsonian SWFs are already widely used in SWF scholarship. They are attractive for measure-theoretic reasons[31] and because of their simplicity. They also correspond to the well-known Atkinsonian index of *inequality*, which is one of the most widely used such indices.

To this point, I have focused on the problem of ranking outcomes. But the SWF framework will not be useful, as a guide to policy choice, unless it ultimately helps to rank choices. In the case where the decisionmaker is operating under complete certainty, this is straightforward: she should make the choice with the best outcome in $O$. In the more realistic case where the decisionmaker is operating under uncertainty, deriving a ranking of choices from the ranking of outcomes becomes trickier.

I argue that the prioritarian framework should use *expected utility theory* to derive a ranking of choices from the ranking of outcomes. A given choice $a$ should be assigned an expected utility term $\mathrm{EU}(a) = \sum_{x \in O} \pi_a(x) w(x)$ —where $\pi_a(x)$ is the probability that action $a$ yields outcome $x$, and $w(x)$ is the social value assigned $x$, i.e., $\frac{1}{1-\gamma}\sum_{i=1}^{N} u_i(x)^{1-\gamma}$. This approach, however, is controversial. Although some scholars within the SWF tradition agree that the ranking of actions should be a function of both the possible outcomes of choice and their probabilities, integrated as per the EU formula just provided,[32] others argue for a different approach.[33]

---

than $y$ if the 2nd-worst-off individual in $x$ is better off than the 2nd-worst-off individual in $y$; and if *their* well-being levels are equal, then $x$ is better than $y$ if the 3rd-worst-off individual in $x$ is better off than the 3rd-worst-off individual in $y$; and so forth.

31 These are the only prioritarian SWFs which are invariant to multiplication of individual utilities by a common positive constant. More precisely, Atkinsonian SWFs are thus invariant if all utilities are nonnegative. No prioritarian SWF is invariant to multiplication of utilities by a common positive constant once negative individual utilities are allowed into the domain of utility vectors. And, quite apart from the invariance issue, the Atkinsonian SWF is not suitable for use with negative utilities because—with negative utilities—the function $g(u_i(x)) = (\frac{1}{1-\gamma})u_i^{1-\gamma}$ is either undefined or, if defined, not both strictly increasing and strictly concave. *See* Adler, *Risk Equity*, *supra* note 1, at 41 n.122.

32 *See, e.g.*, Peter J. Hammond, *Utilitarianism, Uncertainty and Information*, *in* UTILITARIANISM AND BEYOND 85, 90–96 (Amartya Sen & Bernard Williams eds., 1994); Rabinowicz, *supra* note 21; Marc Fleurbaey, *Assessing Risky Social Situations* (Society for Social Choice and Welfare, Working Paper, Feb. 2008), http://www.accessecon.com/pubs/SCW2008/SCW2008-08-00028S.pdf.

33 *See, e.g.*, Elchanan Ben-Porath et al., *On the Measurement of Inequality under Uncertainty*, 75 J. ECON. THEORY 194 (1997); Peter A. Diamond, *Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility: Comment*, 75 J. POL. ECON. 765 (1967); Larry G. Epstein & Uzi Segal, *Quadratic Social Welfare Functions*, 100 J. POL. ECON. 691 (1992). The

For purposes of this Article, I can ignore this dispute. The problem of future generations, to which we now turn, concerns the outcome ranking. How shall we shall take account of future people, possibly nonexistent people, variations in population size, and infinite futures in ranking the outcome set $O = \{x, y, \ldots\}$?

## *II. The Core Case: A Fixed Population and a Finite Future*

There is a substantial body of scholarship in economics that employs some kind of SWF to evaluate policies with intertemporal impacts. Within this body of scholarship, a discount factor is often applied to individual utilities, with the amount of the discount increasing as utilities occur further in the future.[34]

Is this appropriate?[35] To begin, we should distinguish between the propriety of discounting within cost-benefit analysis (where a discount factor is often applied to willingness to pay/accept amounts), and the propriety of discounting within the SWF framework. The two cases raise somewhat different issues, and I will not discuss cost-benefit analysis here.[36] My focus is on SWFs—specifically, on prioritarian SWFs.

We should also distinguish between *intra*personal and *inter*personal discounting.[37] The SWF framework begins with a utility function, which represents the (complete and convergent[38]) extended preferences of fully informed, fully rational, self-interested individuals contemplating the prospect of living different life histories. That

---

dispute about the appropriate application of a social welfare function under conditions of uncertainty is fully reviewed in ADLER, WELL-BEING AND EQUITY, *supra* note 1.

34 *See, e.g.*, CHARLES BLACKORBY ET AL., POPULATION ISSUES IN SOCIAL CHOICE THEORY, WELFARE ECONOMICS, AND ETHICS 253–71 (2005); John Creedy & Ross Guest, *Discounting and the Time Preference Rate*, 84 ECON. REC. 109 (2008); Geoffrey Heal, *Discounting: A Review of the Basic Economics*, 74 U. CHI. L. REV. 59 (2007).

35 There is a vast literature about discounting. *See, e.g.*, Weisbach & Sunstein, *supra* note 12, at 3 n.7.

36 For a discussion of discounting and cost-benefit analysis, see ADLER & POSNER, *supra* note 18, at 173–77. One argument for discounting in the context of cost-benefit analysis—that it is a way to compensate for the fact that future generations will be richer and have a lower marginal utility of consumption—is unavailing in the SWF context, because the marginal utility of consumption is already reflected in the utility function $u(x)$. *See infra* text accompanying notes 46–48.

37 Ricky Revesz draws a similar distinction. *See* Richard L. Revesz, *Environmental Regulation, Cost-Benefit Analysis, and the Discounting of Human Lives*, 99 COLUM. L. REV. 941, 948 (1999) (arguing that "discounting raises analytically distinct issues in the cases of latent harms and harms to future generations").

38 As already explained, I am assuming the completeness and convergence of extended preferences in this Article to simplify the analysis. *See supra* text accompanying note 18.

function will assign each outcome $x$ a *lifetime* utility number $u_i(x) = u(x; i)$, representing the lifetime well-being associated with individual $i$'s life history in outcome $x$. What $u(x; i)$ is will depend on how outcomes are specified (whether in terms of health, consumption, public goods, happiness, other well-being items, or some combination), and on which of the specified characteristics individual $i$ possesses, at various points in his life, in outcome $x$.

A discount factor *may* come into play at this level: at the level of mapping an individual's characteristics in an outcome onto her lifetime utility. For example, imagine that outcomes characterize individuals' life spans and annual consumption for each year alive. So life history $(x; i) = (c^i_{b(x;i)}, c^i_{b(x;i)+1}, \ldots, c^i_{b(x;i)+l(x;i)-1})$ —where $c^i_t$ is the consumption[39] of individual $i$ in year $t$ in outcome $x$; $b(x; i)$ is the year that individual $i$ is born in outcome $x$; and $l(x; i)$ is the number of years that individual $i$ is alive in outcome $x$. Further, let us make the standard assumption that we can assign a "subutility" to an individual's consumption in each time period (here, each year)—for example, $v(c^i_t) = \log c^i_t$ —and that the lifetime utility of a consumption sequence is an additive function of the subutilities.

One possibility is that the additive function is the straight, undiscounted sum of consumption subutility each year. In other words, $u(x;i) = \sum_{t=b(x;i)}^{b(x;i)+l(x;i)-1} v(c^i_t)$. Another possibility is that this additive function incorporates a discount factor, so that consumption events that occur earlier in time are given greater weight in determining lifetime utility. In other words, $u(x;i) = \sum_{t=b(x;i)}^{b(x;i)+l(x;i)-1} G(t)v(c^i_t)$ —where $G(t)$ is a discount factor that decreases as $t$ increases. This latter formula would be an example of what I am calling "intrapersonal discounting." It incorporates a discount factor in generating a lifetime utility number for a given life history from the individual's characteristics in each period (in this example, the individual's annual consumption amounts, as measured by an annual subutility number $v(c^i_t)$ ).

The propriety of intrapersonal discounting raises difficult issues about the nature of well-being that I will not attempt to grapple with here.[40] The traditional philosophical wisdom is that it is irrational to care whether pains and pleasures, consumption events, and other con-

---

39 This could be a vector of consumptions of different goods or a single consumption amount.

40 *See* ADLER, WELL-BEING AND EQUITY, *supra* note 1.

stituents of well-being occur earlier or later in time. Derek Parfit has challenged this traditional view.[41] Assuming it is rationally permissible to care about the temporal position of well-being constituents, it becomes an empirical question (within my framework for assigning utilities) whether fully informed, fully rational, self-interested individuals actually do have this sort of preference.

The sort of discounting that I will challenge in this Article is *interpersonal* discounting: discounting lifetime utilities. This sort of discounting occurs *after* each outcome has been mapped onto a vector of lifetime utilities, representing the lifetime well-being of each individual in the outcome.

Consider, specifically, the "core case" in which the same $N$ individuals exist in all outcomes. By "exist," I mean that the individual is a member of the current generation *or* a past or future generation. She lives now, was alive in the past, or will be alive in the future. In other words, $N$ is the number of the world's *intertemporal population*, which in the core case is assumed to be fixed.

The prioritarian SWF without interpersonal discounting assigns each outcome a social welfare value $w(x)$ equaling $\sum_{i=1}^{N} g(u_i(x))$. In the core case, I believe, this simple, undiscounted formula is most defensible. By contrast, the prioritarian SWF *with* interpersonal discounting takes the form $\sum_{i=1}^{N} D(b(x;i))g(u_i(x))$—where $b(x;i)$ is the date at which individual $i$ is born in outcome $x$, and $D(t)$ is a discount factor which is a decreasing function of time $t$.[42] The upshot of incorporating this interpersonal discount factor into the prioritarian formula is to give less weight, in the social calculus, to the lifetime utilities of future generations as opposed to the current generation—and less weight to the lifetime utilities of individuals who live in the distant rather than near future.

What is wrong with interpersonal discounting? A very basic question, here, concerns the function of the SWF framework. The

---

41 *See* DEREK PARFIT, REASONS AND PERSONS 117–95 (1984).

42 For a full discussion of formulas for social choice that take account of individuals' lifetime utility, birth dates, and length of life, see BLACKORBY ET AL., *supra* note 34, at 253–71.

A variation on this formula puts the discount rate inside rather than outside the $g$-function, i.e., $\sum_{i=1}^{N} g(D(b(x;i))u_i(x))$. Other variations make the discount factor a function of some date other than the individual's birth date, e.g., his death date, or the date midway between birth and death. My arguments against interpersonal discounting apply to all these variations.

framework might be seen as, fundamentally, descriptive: a way to describe society's actual policy preferences. On this view, it is the no-discounting approach which is problematic, because governments in practice are far from neutral between the interests of their current citizens and the interests of future generations.

Whatever their potential descriptive role, SWFs are also a useful *prescriptive* tool. In particular, I defend the SWF framework as a useful tool for *moral* deliberation. It usefully systematizes moral deliberation of a certain sort, namely moral deliberation that starts with welfarism and consequentialism as basic premises. My interest, throughout this Article, is in the appropriate structure of SWFs *where used as tools for moral deliberation*. And my claim in this Part is that interpersonal discounting is inappropriate in that context.

Why? Moral deliberation is *impartial*. Interpersonal discounting violates this impartiality. It gives less weight to the interests of certain individuals, by virtue of a characteristic that is very hard to see as having moral relevance—namely, being born later rather than earlier in the history of the world.

Indeed, given welfarism, we can sharpen the critique of interpersonal discounting. Interpersonal discounting violates *welfarism*. Welfarism says that the ranking of outcomes should be solely a function of individual well-being—with no attention to non-well-being characteristics of individuals.

Welfarism can be formally expressed, in part, through the anonymity axiom. If the ($N$-entry) vector of individuals' lifetime utilities in $x$ is just a permutation of the vector of individuals' lifetime utilities in $y$, $x$ and $y$ should be ranked as equally good. Interpersonal discounting violates the anonymity axiom, by making the ranking of outcomes a function both of individual well-being, and of the point in time when individuals are born. For a simple example, consider a case where $N = 2$. One individual is born earlier, the other later. In outcome $x$, the earlier-born individual has lifetime well-being 50 and the later-born individual has lifetime well-being 100. In outcome $y$, these well-being numbers are switched: the earlier-born individual has lifetime well-being 100 and the later-born individual has lifetime well-being 50. Interpersonal discounting means that $y$ is ranked higher than $x$, which violates anonymity.

It might be objected that anonymity is too demanding a statement of welfarism. I do not see why it is—it seems to capture the basic idea that well-being, and nothing else, should drive the ranking of outcomes—but in any event interpersonal discounting also violates a

different principle that is generally accepted as the very core of welfarism: the Pareto-indifference principle. Pareto indifference says that if each individual is just as well off in outcome *x* as in outcome *y*, the two outcomes are equally good. But interpersonal discounting can violate Pareto indifference.[43] Imagine that each of the *N* individuals achieves the very same level of lifetime utility in *x* as he does in *y*. However, one individual, Jim, has a different birth date in *x* than in *y*. (The objection that this is impossible—that a person's birth date is an essential characteristic of him—fails. For example, if Jim develops from a union of a particular sperm and egg, then he remains the same particular person whether he is born by natural childbirth, or cryogenically frozen as a fertilized ovum by his parents and born 60 years later.[44]) In such a case, using an SWF with interpersonal discounting will rank *x* above *y* or vice versa, but Pareto indifference requires that they be ranked as equally good.

If you like, the argument against interpersonal discounting can be framed in terms of prioritarianism specifically, rather than welfarism generally. Prioritarianism (as I explicate it) sees the ranking of outcomes as emerging from individuals' claims. Consider two individuals, one of whom is better off in *x*, the other better off in *y*. The individuals are born on different dates. It is fair (at least plausibly) that the later-born individual should have a weaker claim between the outcomes if she is at a higher lifetime well-being level in both outcomes than the earlier-born individual. It is also fair (at least plausibly) that the later-born individual should have a weaker claim between the outcomes if her well-being difference between them is less than the earlier-born individual's. But to count the later-born individual as having a weaker claim merely because she is born at a later date is very hard to see as fair. What justifies this deflation of her claim?

In short: the impartial nature of moral deliberation generally, the axioms of welfarism more specifically, and the concern for a fair aggregation of individual claims characteristic of prioritarianism yet more specifically, all cut against interpersonal discounting. In response, it might be argued that this analysis implicitly assumes moral deliberation to adopt an impartial stance vis-à-vis the world's in-

---

43 *See* Tyler Cowen, *Consequentialism Implies a Zero Rate of Intergenerational Discount*, *in* JUSTICE BETWEEN AGE GROUPS AND GENERATIONS 162 (Peter Laslett & James S. Fishkin eds., 1992); JOHN BROOME, WEIGHING LIVES 127 (2004); BLACKORBY ET AL., *supra* note 34, at 253–59.

44 *See* Cowen, *supra* note 43, at 164.

tertemporal population. Why can we not engage in a kind of moral deliberation—if you like, call it "shmoral" deliberation—which focuses on the well-being of a subset of the world's intertemporal population?

Whether we call it "moral" or "shmoral," deliberation that strives to be impartial vis-à-vis some proper subset of the world's intertemporal population—for example, all persons currently existing, or all U.S. citizens, past, present, or future—is certainly a possible enterprise. It is also certainly possible to structure this kind of deliberation using a prioritarian SWF. But the upshot would be an SWF that gives *zero* weight to the lifetime utilities of individuals outside the community of interest—not one that incorporates a discount factor.[45]

What about various standard arguments for discounting?[46] One such argument appeals to economic growth and the declining marginal utility of consumption. Future generations will be at higher consumption levels; increments in their consumption will make a smaller difference to their well-being than increments in the consumption of the current generation; and policy analysis should incorporate a discount factor to reflect that.

Another standard argument appeals to opportunity costs. Given intertemporal financial markets, current resources can be invested at some positive interest rate, yielding more resources in the future. Imagine that the available annual rate is $r$. Imagine, now, that we are considering a regulatory policy that will impose a current reduction in consumption, $\Delta c$, with consumption benefits in twenty years, for the next generation, equaling $\Delta b > \Delta c$. Imagine, further, that the increase in the lifetime utility of the future individuals benefited by the policy is greater than the decrease in the lifetime utility of the current individuals who are made worse off by the policy. If so, a prioritarian SWF without interpersonal discounting might well approve the policy.[47] But the policy may well be a bad idea. In particular, imagine that $\Delta c(1+r)^{20} > \Delta b$. In that case, we could produce greater consumption and, therewith, utility for the future generation, with the very

---

45 To be sure, the deliberator who wanted to be impartial vis-à-vis some proper subset of the world's population *would* presumably want to be sensitive to the fact that harms and benefits to persons outside the community of interest might, in turn, affect the well-being of individuals inside the community of interest—for example, because these insiders care about the outsiders. However, such impacts would show up in the utility function of the insiders.

46 *See* ADLER & POSNER, *supra* note 18, at 173–77.

47 Whether the prioritarian SWF would, in fact, approve the policy would depend on the utility levels of the future and present individuals.

same consumption and utility cost for the present generation, by investing the resources in the market at interest rate *r*.

A final argument appeals to uncertainty. In general, we are less certain about what the impact of a policy will be on unborn individuals than on the current generation.

These three standard arguments all make vital points, but none justify interpersonal discounting. A prioritarian SWF with the undiscounted form $\sum_{i=1}^{N} g(u_i(x))$ is *already* fully sensitive to considerations of economic growth/declining marginal utility, opportunity costs, and uncertainty.

The first point is handled by the mapping from the outcome ranking to the choice ranking, and by the individual utility function $u(.)$. If, given one or another choice in the choice set, it is likely that future generations will be richer (at a higher consumption level) than the current generation, then this just means that the subset of outcomes in which future generations are richer than the current generation will be assigned a high probability (conditional on that choice),[48] and that the subset of outcomes in which future generations are poorer than the current generation will be assigned a low probability (conditional on that choice). And if consumption does indeed have declining marginal utility, that will be reflected in the function $u(.)$ that maps outcomes (characterized in part in terms of individual consumption) onto individual lifetime utility.

Opportunity costs are most straightforwardly handled within the SWF framework by representing the option of investing current resources as an additional policy choice. Where intertemporal markets exist, the choice set considered by policy analysts should not merely include $\{sq, a, b, \ldots\}$, that is, the status quo option of inaction and

---

48 I assume here that the ranking of actions in the choice set $A$ is a function of probability numbers expressing the probability that a given action $a$ in this set would yield a given outcome $x$ in the outcome set $O$—probability numbers of the form $\pi_a(x)$—as well as the utility function $u(.)$ that maps each outcome onto a utility vector, and the SWF $\sum_{i=1}^{N} g(u_i(x))$ that produces a ranking of those vectors. The probability term $\pi_a(x)$ is what I am calling, roughly, the probability of an outcome "conditional" on a choice. Whether $\pi_a(x)$ is, strictly, a conditional probability implicates the debate between so-called "causal" and "evidential" decision theory, which I need not further discuss here. *See* ADLER, WELL-BEING AND EQUITY, *supra* note 1. Further, my assumption that the ranking of actions in the choice set is a function of the $\pi_a(x)$ values, the utility function, and the SWF is more generic than the specific claim that this information is integrated via the EU formula—an issue I leave open in this Article. *See supra* text accompanying notes 32–33.

various regulatory policies, but also {*sq*, *a*, *b*, . . . , *inv*}, where *inv* is the option of investing resources in intertemporal markets at a positive interest rate.

Finally, and in line with decision theory more generally, the SWF framework straightforwardly handles uncertainty at the level of mapping actions to outcomes. Uncertainty is a feature of a particular decisionmaker, poised to choose from some choice set. It shows up in her probability distribution, conditional on each choice in that set, over the outcome set.[49] Uncertainty should not (and need not) be handled by distorting the outcome ranking itself, for example by incorporating an interpersonal discount factor into the formula used to produce the ranking.[50]

Let me conclude by returning to the point that an SWF without interpersonal discounting might argue for policies that are dramatically different from current policies. This point might be framed, not as a challenge to the no-discounting approach understood as a purported description of actual practice—again, my interests are prescriptive, not descriptive—but indeed as a challenge to the *prescriptive* credentials of the no-discounting approach.

The point might be articulated as follows. "No-interpersonal-discounting requires policies that are very different from our current policies. Further, it is counterintuitive that the current generation should pursue these new policies. Moral deliberation is a matter of striving for reflective equilibrium, by giving weight both to general principles and to intuitions about particular cases. No-interpersonal-discounting, in requiring a radical revision of current policies, is *morally* problematic because it fails the test of reflective equilibrium."

---

49 *See supra* note 48.

50 One variation on the uncertainty argument for discounting points to an extinction risk. *See* STERN, *supra* note 10, at 50–54; Antoine Bommier & Stephane Zuber, *Can Preferences for Catastrophe Avoidance Reconcile Social Discounting with Intergenerational Equity?*, 31 SOC. CHOICE & WELFARE 415 (2008); Partha Dasgupta, *Discounting Climate Change*, 37 J. RISK & UNCERTAINTY 141, 160 (2008). This argument moves outside the core case. Instead, the date after which individuals no longer exist, and the number of individuals who exist, can vary across outcomes.

If the existence of a given generation is likelier than succeeding generations, and less likely than preceding generations, then an argument for ranking outcomes with a discount factor to reflect this extinction risk arises. However, a more transparent and flexible approach to taking account of extinction risk would seem to be ranking outcomes without a discount factor, and incorporating the risk into the representation of each action as a probability distribution across outcomes. The flexibility of this approach would be particularly advantageous, it would seem, if we wish to allow the extinction risk to vary depending on which policy is undertaken (e.g., whether or not we take steps to mitigate a catastrophic event), rather than being exogenous.

In general, the fact that a moral theory requires a radical revision of current policies does not entail that the theory is counterintuitive in the reflective equilibrium sense—even to those engaging in the policies.[51] However, no-interpersonal-discounting has a potential result that *is* both revisionary and morally counterintuitive.

The worry, specifically, is that a no-interpersonal-discounting approach, together with a positive interest rate in intertemporal markets, might require the current generation to impoverish itself for the future.[52] To see this worry in a highly simplified case, imagine that: the world has two periods; individuals live for a single period; there are an equal number of individuals who are alive in each period; and an individual's lifetime well-being is just a linear function of his consumption during the one period he is alive, i.e., $u_i(x) = v(c_i) = kc_i$, where $c_i$ is the amount that $i$ consumes in outcome $x$ during the period he is alive. Consumption comes from a store of collective resources. Individuals who consume nothing have utility zero.[53] The leaders of the earlier generation must decide whether to allow that generation to consume all of the resources, or invest some fraction at an interest rate $r > 0$, for consumption by the later generation. If the leaders make this decision by employing a utilitarian SWF with no interpersonal discounting, they reach the morally counterintuitive conclusion that they should invest the entire stock of resources, leaving none for the current generation.

This conclusion *is* counterintuitive, but it can be avoided by revising features of the SWF framework *other* than no-discounting—which, as argued, is a feature we should be particularly unwilling to relinquish in reflective equilibrium, because it is tied to impartiality. First, using a utility function $v(c_i)$, which is not a linear function of consumption, but rather has the plausible property that the marginal utility of consumption declines as consumption increases, will reduce the amount of resources that the earlier generation is required to invest for the future.

Second, and independently, shifting from utilitarianism to prioritarianism helps to avoid the counterintuitive conclusion. Consider

---

51 There certainly can be societies that are immoral in some fundamental sense, e.g., racist, and are also seen by some of the members of these societies to be immoral.

52 *See* Geoffrey Brennan, *Discounting the Future, Yet Again*, 6 POL., PHIL. & ECON. 259, 268 (2007).

53 This premise is to simplify the analysis, and might mean that without consumption individuals die prematurely and end up with a life no better than nonexistence; or that they subsist but end up with a life no better than nonexistence.

the Atkinsonian SWF $\frac{1}{1-\gamma}\sum_{i=1}^{N} u_i(x)^{1-\gamma}$. Assume, for simplicity, that $v(c_i) = kc_i$. As already mentioned, with $\gamma = 0$, the SWF is utilitarian. As $\gamma$ increases, the SWF becomes increasingly inequality-averse[54] and, at the limit, approaches leximin.[55] Note now that if $\gamma$ is greater than zero, the current generation will never be required to give all of the resources for the future. (Intuitively, by investing a given increment of the resources, the current generation produces a greater well-being change for the future generation than if it had consumed the increment; but the current generation also lowers its well-being level, thereby giving itself a stronger claim to the next increment.) Further, the greater the value of $\gamma$, the less the current generation is required to invest. Finally, at the limit, with a leximin social ordering, the current generation is required to invest only so much as to produce the very same consumption amounts for both generations.[56]

In short, cases in which a no-interpersonal-discounting rule plus a positive interest rate lead to a certain degree of well-being disparity

---

54 Take any two individuals, one better off, one worse off. Consider the ratio between the change in the value of the SWF produced by giving a small increment of utility to the worse-off individual, and the change produced by giving a small increment to the better-off individual. That ratio increases without limit as $\gamma$ increases.

55 *See* ADLER, WELL-BEING AND EQUITY, *supra* note 1; Adler, *Risk Equity*, *supra* note 1, at 40–45.

56 To show this formally, let us denote by $S$ the stock of resources held by the current generation, and $N$ the number of individuals in each generation. The leader of the current generation chooses some amount $c$ for the first-generation individuals to consume. With an interest rate of $r$, this yields $(S\text{-}c)(1+r)$ for the future generation. If the leader makes this decision using the Atkinsonian SWF, without discounting the future generation's utility, she chooses $c$ so as to maximize $\frac{1}{1-\gamma}(N(c/N)^{1-\gamma} + N[(S-c)(1+r)/N]^{1-\gamma})$. (To see why this is the correct formula, note that, if a given amount of total consumption is allocated to some generation, the SWF is maximized by spreading it evenly among the members of that generation.) This formula is maximized where $c$ has the value $\frac{S}{(1+r)^{\frac{1-\gamma}{\gamma}}+1}$. The derivative of this expression with respect to $\gamma$, the Atkinsonian inequality-aversion parameter, is positive—which means that as $\gamma$ increases, optimal first-generation consumption increases (and second-generation consumption decreases). In other words, by increasing the degree of inequality-aversion, the degree to which the first generation is required to impoverish itself for the second decreases. Note also that, as $\gamma$ approaches zero, so that the SWF becomes utilitarian, the optimal value of $c$ approaches zero. Conversely, as $\gamma$ approaches infinity, so that the SWF becomes a leximin SWF (giving absolute priority to the worst-off individual), the optimal value of $c$ approaches $\frac{S(1+r)}{2+r}$. At this value of $c$, the amount consumed by the second generation is $(S\text{-}c)(1+r)$, which equals $\frac{S(1+r)}{2+r}$ — confirming that, as the SWF approaches leximin, the two generations approach a point at which they consume equal amounts.

between earlier and later generations should be seen as cases that help us calibrate the degree of inequality aversion of our SWF. These cases—like synchronic examples in which some individuals are made worse off than others for the sake of an increase in overall welfare—help us to determine, as a matter of moral deliberation, how much we morally value overall welfare versus welfare equality. They help us to reach a point of reflective equilibrium on the spectrum between utilitarianism and leximin. They need not and should not prompt us to take the radical step of departing from welfarism and moral impartiality by adopting interpersonal discounting.

## *III. Beyond the Core Case: Non-Identity, Variable Population, and Infinity*

In the core case of a fixed and finite population, the treatment of future generations is straightforward—or so I have just argued. Use the simple prioritarian formula $\sum_{i=1}^{N} g(u_i(x))$ —without a discount factor attached to individuals' lifetime utilities, to rank outcomes.

When we depart from the core case, matters become murkier.

### *A. Non-Identity*

A large literature in moral philosophy addresses "non-identity" cases: where some action causes the very existence of some individual who would otherwise not exist, rather than merely harming or benefiting some individual who exists independent of the action.[57] Gregory Kavka nicely summarizes the problem, focusing particularly on the implications for future generations.

> I have heard a rumor, from a reliable source, that I was conceived in New Brunswick, New Jersey. Had my father been on duty at Camp Kilmer that fateful weekend, or had there been an earthquake in central New Jersey at the wrong moment, or had any of innumerable other possible events oc-

[57] *See, e.g.*, JOHN BROOME, *The Value of a Person*, *in* ETHICS OUT OF ECONOMICS 228 (1999) [hereinafter BROOME, *Value*]; BROOME, *supra* note 43; DAVID HEYD, GENETHICS (1992); PARFIT, *supra* note 41, at 351–79; MELINDA A. ROBERTS, CHILD VERSUS CHILDMAKER (1998); Krister Bykvist, *The Benefits of Coming into Existence*, 135 PHIL. STUD. 335 (2007); Caspar Hare, *Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?*, 117 ETHICS 498 (2007); Nils Holtug, *On the Value of Coming into Existence*, 5 J. ETHICS 361 (2001); Gregory S. Kavka, *The Paradox of Future Individuals*, 11 PHIL. & PUB. AFF. 93, 93–94 (1981); Jan Narverson, *Utilitarianism and New Generations*, 76 MIND 62 (1967); Josh Parsons, *Axiological Actualism*, 80 AUSTRALASIAN J. PHIL. 137 (2002); James Woodward, *The Non-Identity Problem*, 96 ETHICS 804 (1986).

curred, the particular sperm and egg cells from which I developed would never have joined, and I would never have existed. This observation about the precariousness of my origin reflects a basic fact about identity and existence that seriously complicates attempts to understand our moral relationship to future generations. Which *particular* future people will exist is highly dependent upon the conditions under which we and our descendants procreate, with the slightest difference in the conditions of conception being sufficient, in a particular case, to insure the creation of a different future person.

This fact forms the basis of a surprising argument . . . to the effect that we have no moral obligation to future generations—beyond, at most, the next few—to promote their well-being. . . . Let us assume that sameness of genetic structure is, for practical purposes, a necessary condition of personal identity. . . . As a result, any proposed policy that would directly or indirectly affect conditions for conception (that is, who mates with whom, and when) on a worldwide scale over a significant period of time would result in an entirely different set of human individuals coming into existence than otherwise would. Now suppose, as seems reasonable, that the various broad-ranging policies designed to promote better living conditions for future generations . . . would, if practiced, affect conditions for conception worldwide. Further, let us allow that if we do not practice these policies, future people will not be so badly off that it would have been better for them never to have existed.

Granted these assumptions, are we obligated to practice controlled growth policies in order to bring about better living conditions for future people? No, for we harm no one if we allow an alternative policy, call it laissez faire. Consider the individuals in the overcrowded world that would result from laissez faire. They are not worse off than if we had acted to bring about the less crowded state of the world, for in that case they would not have existed. And, by hypothesis, their existence is not worse than never having existed. But these people are all the people there *are*, if we practice laissez faire. Thus, in doing so, we make no one worse off (than he otherwise would be) and hence do nothing wrong. We are therefore under no moral obligation to future people to pursue controlled growth policies in order to promote their well-

being.[58]

Much of the literature has discussed non-identity problems from a nonconsequentialist perspective. My focus, here, is on its implications for consequentialism, specifically prioritarianism.

The non-identity problem arises, within prioritarianism, where an outcome set $O$ is such that there are some individuals who exist in some but not all outcomes in the outcome set. For short, let us call these individuals "potential nonexistents."

Note that an outcome set with varying numbers of individuals existing in different outcomes necessarily contains potential nonexistents. However, an outcome set with the same number of individuals in every outcome may also contain potential nonexistents. Let us use $N(x)$ to mean the number of individuals who exist in outcome $x$. It is possible that $N(x) = N(y)$, and yet that that some individuals exist in $x$ but not $y$ and vice versa.[59]

It also bears note that potential nonexistents may not be potential future individuals. Outcomes are simplified descriptions of reality—past, present, and future. It is possible for a decisionmaker at some time $T$ to be using an outcome set that includes outcomes $x$ and $y$, where some individual $i$ is born in $x$ and dies in $x$ at some time prior to $T$, and never exists in $y$. Individual $i$ would be a potential past individual who exists in $x$ but not $y$.

It is certainly true that decisionmakers often have the causal power to bring some potential future individuals into existence, but never have the causal power to bring past or present individuals into existence (because causation runs forward). If, in addition, the decisionmaker knows for certain which individuals are alive at the time of her decision or were alive at some point in the past, then she can narrow down her outcome set to include only outcomes that specify the existence of these individuals. For any pair of outcomes, $x$, $y$, in this restricted outcome set, potential nonexistents will be potential future individuals. In this setup, the non-identity problem *is*, strictly, a problem of "future generations."

However, the difficulty that potential nonexistents pose for the ranking of outcomes, and possible resolutions of that difficulty, do not hinge on whether those potential individuals are born (in outcomes

---

58 Kavka, *supra* note 57, at 93–94 (citations omitted).

59 *See* PARFIT, *supra* note 41, at 355–56 (distinguishing between "Same People Choices" and "Different People Choices" and, within the latter category, between "Same Number Choices" and "Different Number Choices").

where they exist) at a time before or after the time of decision. So my analysis will be more generic.

What is the difficulty? It arises because prioritarians espouse what has been termed the "person affecting principle."[60] There are various formulations of the principle in the literature but, roughly, it says that one outcome is better than another in virtue of being better for individual persons. Outcomes are not morally good or bad in some impersonal sense; rather, the moral goodness of an outcome is derivative of its goodness for persons.

The person-affecting principle flows naturally from the fairness-based account of prioritarianism that I adopt. Morality, on this view, is the set of norms that derives from a desire to fairly and impartially respect the claims—claims to well-being—of different human persons. Indeed, the person-affecting principle is bound up with *welfarism* more generally. If we give up the premise that outcomes are morally better or worse just in virtue of being better or worse for persons, what would justify the insistence that the ranking of outcomes must be solely a function of well-being? Why not, for example, allow considerations quite distinct from the well-being of humans, or even animals, to influence the ranking—for example, the intrinsic value of ecosystems?

But consider, now, a pair of outcomes, $x$ and $y$, such that one individual, Jill, exists in $x$ but not $y$. It would seem that $y$ is neither worse for Jill than $x$, nor better for Jill than $x$. Why? Surely there is a conceptual connection between "worse for" and "worse off than." Outcome $w$ is worse for Steve than outcome $z$ iff: were outcome $w$ to be the actual outcome, Steve would be *worse off* than he would have been if outcome $z$ had been the actual outcome. But: were outcome $y$ to be the actual outcome, Jill would not have the property of being worse off, or better off, than she would have been had $x$ been the actual outcome, because Jill does not exist in outcome $y$.

The conceptual connection between "worse for" and "worse off than" also means, it seems, that the outcome $x$ in which Jill exists is neither better for her than nonexistence, nor worse for her than nonexistence. Why? It is not the case that Jill, in outcome $x$, is better off or worse off than she would have been in outcome $y$, because she would not have existed if outcome $y$ had been the case. Lest this ob-

---

60 For a recent discussion of the person-affecting principle with reference to prioritarianism, see Holtug, *supra* note 21. Seminal discussions of the principle and its implications for population policy include Narveson's and Parfit's. *See* PARFIT, *supra* note 41, at 391–417; Narveson, *supra* note 57.

servation not persuade, a different route to the same conclusion is that *x* cannot be better or worse for Jill than *y*, because *x* is better for Jill than *y* only if *y* is worse for Jill than *x*, and *x* is worse for Jill than *y* only if *y* is better for Jill than *x*. But the previous paragraph established, seemingly, that *y* is neither better nor worse for Jill than *x*.

We have reasoned ourselves to the proposition that outcomes in which potential nonexistents exist are neither better nor worse for them than outcomes in which they do not exist, nor is nonexistence better or worse for them than outcomes in which they do exist. This proposition, together with the person-affecting principle, implies that a potential nonexistent should be ignored in comparing an outcome in which she exists to one in which she does not.[61] In other words, we have reached the conclusion that the interests of potential nonexistents should be *discounted to zero* in comparing such pairs of outcomes. Grappling with that unpalatable conclusion is the nub of the "non-identity problem," as I see it, in the context of prioritarianism.

One response is to accept the conclusion. The ranking of outcomes *x* and *y* should be just a function of the well-being of individuals who exist in both outcomes. This conclusion, however, is intuitively unpalatable. Jill, who exists in *x* but not *y*, might have a very good life indeed in *x*. Should that not count as a reason in favor of *x*? Alternatively, Jill's life in *x* might be hellish—a life of unremitting pain. Should that not count as a reason against *x*?

Accepting the conclusion also yields a grave formal difficulty. There is very strong reason to think that the ranking of outcomes generated by any plausible consequentialist view should be *transitive*. If *x* is morally better than *y*, and *y* morally better than *z*, then *x* is morally better than *z*. But ranking pairs of outcomes by ignoring individuals who exist in one but not both outcomes yields intransitivities, as Table 1 shows:[62]

**Table 1. Potential Nonexistents and Intransitivity**

| | **Outcomes** | | |
|---|---|---|---|
| **Individuals** | *x* | *y* | *z* |
| **Henry** | 5 | 6 | 4 |
| **Jane** | NE | 1 | 4 |
| **Sally** | 10 | 10 | 10 |

61 Note that we have reasoned ourselves to this conclusion independent of whether potential nonexistents have very good or very bad lives in outcomes where they exist.

62 *See* BROOME, *Value*, *supra* note 57.

Jane does not exist in $x$. Any prioritarian SWF will rank the vector of utilities (6,10) over (5,10); will rank the vector of utilities (4,4,10) over (6,1,10); and will rank the vector of utilities (5,10) over (4,10). So if we ignore Jane in comparing outcomes $x$ and $y$, and in comparing outcomes $x$ and $z$, we produce the intransitivity $y$ better than $x$, $z$ better than $y$, and $x$ better than $z$.

A different response to the problem of potential nonexistents is to reject the person-affecting principle. Where Jill exists in $x$ but not $y$, one might say that facts about her well-being in $x$ create an *impersonal* reason relevant to the ranking of $x$ and $y$. This is the position that Derek Parfit adopts in *Reasons and Persons*.[63] As just discussed, however, rejecting the person-affecting principle is not a comfortable position for prioritarians.

A third possibility is to reject the proposition that as between an outcome $x$ in which a person exists and an outcome $y$ in which she does not, neither outcome is better nor worse for her than the other. I tentatively suggest that this proposition should indeed be rejected—by rejecting the supposed conceptual connection between "worse for" and "worse off than."[64]

In general, a ranking of life histories (as I see it) does *not* involve comparing the properties of an individual in some outcome to the properties of that very same individual in some other outcome. Consider the comparison of life history $(x; i)$ to life history $(y; j)$, where $i$ and $j$ are distinct individuals. That comparison is undertaken (as I see it) by asking whether fully informed, fully rational, self-interested individuals would prefer to be $i$-in-$x$ or $j$-in-$y$. It does not depend upon, or imply, the premise that one particular person is better or worse off in $x$ than he would have been in $y$, or has or lacks some property in $x$ that he lacks or has in $y$.

Similarly, a comparison of a life history $(x; i)$ to nonexistence does not depend upon, or imply, the premise that some particular individual $(i)$ is better off existing than he would have been not existing. Rather, it involves asking whether fully-informed, fully-rational, self-interested individuals[65] would extendedly prefer to be $i$-in-$x$ rather

---

63 PARFIT, *supra* note 41, at 378, 447.

64 Nils Holtug argues, in a different way, that an individual can be better or worse off existing or not existing, and therefore that non-identity problems do not jeopardize the person-affecting principle. *See* Holtug, *supra* note 57; Holtug, *supra* note 21; Nils Holtug, *Utility, Priority and Possible People*, 11 UTILITAS 16 (1999) [hereinafter Holtug, *Utility*].

65 To be sure, where an outcome set involves potential nonexistents, one needs to ask: whose extended preferences should we look to in comparing life histories? In the core case, one

than not existing at all. To say that $(x; i)$ is *better for i* than nonexistence is just to say that the answer to this question is affirmative. To say that $(x; i)$ is *worse for i* than nonexistence is to say that its answer is negative.

It might be objected that it is incoherent for someone to ask herself whether she has an extended preference in favor of a given life history, as compared to nonexistence. But this question seems perfectly intelligible. Indeed, the thought experiment of comparing life histories to nonexistence is a central part of the prioritarian framework, as I see it. Even in the "core case" of a fixed and finite population, we need a "zero point" to help calibrate the utility scale—and we most readily do that by considering whether individuals would prefer or disprefer various life histories to nonexistence.[66]

Fully elaborating the analysis sketched here would mean grappling with subtle issues involving existence, modality, personhood, properties, and extended preferences—an enterprise I cannot undertake at greater length in this Article. The analysis might well founder on closer inspection. Still, I tentatively embrace the following position. Existence can be better or worse for an individual than nonexistence. Nonexistence can be better or worse for an individual than existence. Where an outcome set contains potential nonexistents, their interests should be taken into account by assigning them a utility level of zero in the outcomes where they do not exist.

Assigning them a utility level of zero, and then doing what? Imagine, first, that an outcome set contains potential nonexistents, but the same number $N$ of individuals exist in all outcomes. In that case, the answer is straightforward. Use the formula $\sum_{i=1}^{N} g(u_i(x))$ to rank the outcomes.

Imagine, next, that the outcome set contains potential nonexistents, and that the number of individuals who exist in each outcome is not the same. This brings us to the thorny issue of variation in population size, which I will discuss below.

Two other responses to the non-identity problem bear mention. One response, suggested by Partha Dasgupta, is to make the ranking

---

looks to the extended preferences of the $N$ individuals. What happens outside the core case? I lack space to pursue that question here.

66 John Broome denies that a life history can be better or worse for a person than nonexistence. He therefore pursues a different approach to fixing the zero level, which identifies a neutral level for continuing to live, and then defines a neutral life as one which is constantly at this neutral level. *See* BROOME, *supra* note 43, at 66–68, 233–35, 241–53.

of outcomes *relative* to a population. For each outcome $x$, we can rank the entire set $O$ of outcomes relative to the individuals who exist in $x$.[67] A difficulty here is that this approach may end up giving us very little moral guidance. At the end of the day, to decide what to do, we need a single ranking of the outcome set, not just a set of relative rankings.

Another response, suggested by Josh Parsons, is "actualism." Roughly, we compare $x$ and $y$ by looking to the interests of individuals who have actually existed, or will actually exist. Assuming determinism, the actual past *and future* population of the universe is now fixed. Under "actualism," it is *their* well-being that drives the ranking of all outcomes.[68] This approach, however, produces deep paradoxes, as Caspar Hare has shown.[69]

### *B. Variation in Population Size*

Assume that we assign individuals who exist in some but not all outcomes a utility of zero in the outcomes where they do not exist.[70] In cases where the population number is variable, we *could* rank outcomes using the following prioritarian formula: outcome $x$ is as least as good as outcome $y$ iff $\sum_{i=1}^{N(x)} g(u_i(x)) \geq \sum_{i=1}^{N(y)} g(u_i(y))$, where $N(x)$ is the number of individuals who exist in $x$ and $N(y)$ the number of individuals who exist in $y$. Call this the "total" prioritarian approach. Unfortunately, it runs smack into what Derek Parfit calls the "repugnant conclusion."[71]

The repugnant conclusion is often presented as a problem for *utilitarianism*. Imagine an outcome $x$ in which there are $N$ individuals.

---

67 *See* PARTHA DASGUPTA, AN INQUIRY INTO WELL-BEING AND DESTITUTION 377–97 (1993). Dasgupta's approach is actually somewhat more complicated than simply ranking the outcomes relative to each population. For a clear description and critical discussion, see BROOME, *Value*, *supra* note 57; BROOME, *supra* note 43, at 157–63.

68 *See* Parsons, *supra* note 57.

69 *See* Hare, *supra* note 57.

70 Note that this could happen in two ways: (1) by following the approach I recommend (coupling the person-affecting principle with the position that nonexistence can be better or worse for a person than existence), or (2) by following an impersonal approach. *See supra* text accompanying notes 60–63.

71 *See* PARFIT, *supra* note 41, at 381–90. There is a substantial literature on the repugnant conclusion and other population-size problems. Two recent, authoritative contributions, which also cite much of the prior literature, are BLACKORBY ET AL., *supra* note 34, at 129–208; and BROOME, *supra* note 43. For discussions with specific reference to prioritarianism, see Campbell Brown, *Prioritarianism for Variable Populations*, 134 PHIL. STUD. 325 (2007); Holtug, *Utility*, *supra* note 64; Brown, *supra* note 21, at 187–223.

The average lifetime well-being in $x$ is a very high level $L$. To sharpen the case, imagine that *everyone* in $x$ is at lifetime well-being level $L$. Outcome $x$ might seem like a very good outcome indeed—particularly if $N$ is large. What could be better than to have lots of people at a very high level of well-being?

However, ranking outcomes in accordance with "total" utilitarianism—that is, ranking $x$ as being at least as good as $y$ iff $\sum_{i=1}^{N(x)} u_i(x) \geq \sum_{i=1}^{N(y)} u_i(y)$—has the following unpleasant implication: for any positive level of well-being $L^* < L$, however close to zero $L^*$ might be, there is some outcome $y$ in which everyone is at level $L^*$, and yet $y$ is better than $x$.[72] Consider the outcome $y$ in which there are $M$ individuals, where $M > NL/L^*$. $L^*$ might be a life barely worth living; $L$ might be an absolutely terrific life. And yet ranking outcomes using the formula, $x$ at least as good as $y$ iff $\sum_{i=1}^{N(x)} u_i(x) \geq \sum_{i=1}^{N(y)} u_i(y)$, forces us to the conclusion that $y$ is a better outcome than $x$. "Total" utilitarianism implies the repugnant conclusion that we can always "compensate" for a loss in average individual well-being, all the way down to barely above zero, by a sufficiently large expansion of the population.

The repugnant conclusion is no less a problem for "total" prioritarianism. Assume that there are $N$ individuals who exist in $x$ and who are all at lifetime well-being level $L$ in $x$. Consider again any positive level $L^*$ of lifetime well-being, however close to zero. Then once again there is some outcome $y$ in which everyone is at level $L^*$, and yet $y$ is better than $x$. Indeed, because "total" prioritarianism sums a strictly increasing and concave transformation of individuals' utility, it can be shown that the expansion of the population required to make $y$ better than $x$ is *less* than the number required in the utilitarian case. It is less than $NL/L^*$.[73] Ranking outcomes by simply

---

72 Outcome sets are structured to be responsive to decisionmakers' bounded rationality and thus may well exclude some logically possible outcomes. So, strictly, the worry is that there is some possible outcome set which includes both an $x$ in which everyone lives at level $L$, and a $y$ in which everyone lives at $L^*$, arbitrarily close to zero—and that the "total" formula, applied to this outcome set, ranks $y$ over $x$.

73 More precisely, the claims in this paragraph as well as the remaining paragraphs in this Section are true if the strictly increasing and concave function used by the prioritarian to transform individual utility—the $g(.)$ function in the formula $w(x) = \sum_{i=1}^{N(x)} g(u_i(x))$—is normalized so that $g(0) = 0$. *See* BLACKORBY ET AL., *supra* note 34, at 165 (assuming such normalization in discussion of generalized utilitarianism). In the case of "total" prioritarianism, this normalization means that adding a person with utility zero (a person whose life is as good as nonexistence) to an outcome does not affect the $w$-value of the outcome.

summing a strictly increasing and strictly concave function of individual utility means that the loss in social value in moving an individual from $L$ to $L^*$ is less than proportional to the social gain realized by moving an individual from 0 to $L^*$. This is uncontroversial, for prioritarians, in the core case of a fixed population. But it has the unpleasant consequence, once population size is allowed to vary, that it is even easier for "total" prioritarians to use an expansion of the population as a way to offset losses in average individual well-being than for the "total" utilitarian.

Charles Blackorby, Walter Bossert, and David Donaldson have undertaken an exhaustive, formal analysis of the repugnant conclusion and possible responses within the context of utilitarianism, prioritarianism, and other SWFs.[74] I refer the reader to their work. One possibility, for prioritarians, is to switch from "total" to "average" prioritarianism. Assign each outcome $z$ a value equaling $\frac{1}{N(z)}\sum_{i=1}^{N(z)} g(u_i(x))$ and rank them accordingly. This is, of course, the prioritarian analogue of "average" utilitarianism.

However, "average" prioritarianism has very counterintuitive conclusions. First, it violates what Blackorby, Bossert, and Donaldson call the "negative expansion principle." It should never be possible to improve an outcome by adding individuals with negative utility. But, in the case of "average" prioritarianism, this is possible.[75]

Second, "average" prioritarianism violates a separability-across-persons axiom that I see as part and parcel of prioritarianism.[76] Con-

---

Assume, therefore, that $g(0) = 0$ and that in some outcome $x$ there are $N$ individuals at level $L$. Consider a positive level of utility $L^*$, however small. Let $M = NL/L^*$. $M$ is the "break-even" value for the total utilitarian. The total utilitarian says that an outcome with $M$ individuals at $L^*$ is just as good as $x$; that outcomes with more than $M$ individuals at $L^*$ are better than $x$; and that outcomes with fewer than $M$ individuals at $L^*$ are worse than $x$. Observe, now, that $L^* = (N/M)L$. Because $g(0) = 0$, and because $g(.)$ is strictly increasing and strictly concave, it follows that $g(L^*) > (N/M)g(L)$, with both $g(L^*)$ and $g(L)$ positive.

Rearranging terms, we see that $M > Ng(L)/g(L^*)$. The "break-even" population size $M+$ for the "total" prioritarian—such that an outcome with $M+$ individuals at $L^*$ is just as good as $x$, outcomes with more than $M+$ individuals at $L^*$ are better than $x$, and outcomes with fewer than $M+$ individuals are worse than $x$—is just $Ng(L)/g(L^*)$. So $M+$ is less than $M$.

For similar observations to the effect that the "total" prioritarian finds it even easier than the "total" utilitarian to use population expansion to compensate for reductions in average well-being levels, see Brown, *supra* note 21, at 210–17; Holtug, *Utility*, *supra* note 64, at 32–35.

74 BLACKORBY ET AL., *supra* note 34, at 129–208.

75 *See id.* at 172.

76 Blackorby, Bossert, and Donaldson discuss separability-across-persons under the head-

sider outcomes $x$ and $y$, where Jim is equally well off in $x$ and $y$. Separability-across-persons means that the particular level of Jim's well-being should not influence the ranking of $x$ and $y$. It should not matter whether Jim is at level $L+$ in both outcomes, or instead at level $L++$ in both outcomes, or instead at level $L+++$ in both outcomes, and so forth. After all, whatever Jim's well-being level happens to be, he has no claim in favor of $x$ over $y$, or $y$ over $x$. Whatever that level happens to be, he is equally well off in both outcomes. But "average" prioritarianism can readily violate separability-across-persons.

Blackorby, Bossert, and Donaldson suggest a different response to the repugnant conclusion. Adding an individual to the population should be seen as increasing social welfare iff the individual is above a "critical level": a positive level of well-being $c^*$ which is better than nonexistence.[77] In the case of prioritarianism, the "critical level" formula becomes: $x$ is at least as good as $y$ iff $\sum_{i=1}^{N(x)}[g(u_i(x))-g(c^*)] \geq \sum_{i=1}^{N(y)}[g(u_i(y))-g(c^*)]$. The upshot of this "critical level" approach is that expanding the population by adding lives that are barely worth living—lives which are above zero, but below the critical level—is seen as making the outcome worse. (By contrast, "total" prioritarianism always sees such an addition as improving the outcomes.) The critical level, like other moral parameters,[78] might be set by reflecting on hypothetical cases.

Critical-level prioritarianism is not problem-free. Although it avoids what Blackorby, Bossert, and Donaldson call the "negative expansion principle," it violates another principle involving negative utilities that they call "priority for lives worth living."[79] This says that, for every outcome $x$ in which everyone's well-being is positive, and every outcome $y$ in which everyone's well-being is negative, $x$ must be ranked better than $y$. How can it ever be better to have a world in which everyone is better off not existing, than a world in which everyone is better off existing?

Perhaps the best solution, on balance, is to revert to "total" pri-

---

ing of various "independence" axioms. *See id.* at 159–60.

77 They generalize this formula by contemplating a range of critical levels. *See id.* at 219–21, 248–52. They also consider other sorts of variations, for example "restricted" and "number-sensitive" critical-level views. *See id.* at 136–51, 165–71. John Broome argues for critical-level utilitarianism. *See* BROOME, *supra* note 43, at 254–64.

78 For example, the inequality-aversion parameter in the Atkinsonian SWF.

79 BLACKORBY ET AL., *supra* note 34, at 166 (discussing properties of critical-level generalized utilitarianism, in particular where the critical level is positive).

oritarianism and accept the repugnant conclusion.[80] In $x$, there are $N$ individuals at a high level of well-being $L$. In $y$, those individuals are brought down to a very low positive level $L^*$, but many more individuals exist who do not exist in $x$. Each such individual has a claim to $y$ rather than $x$ (at least if one accepts the "solution" to the non-identity problem I offered in the prior section). Outcome $y$ is better for each such individual than nonexistence, and nonexistence is worse for her—if only by a little bit. Should we not recognize that the number of such claims can become sufficiently large to counterbalance the claims of the $x$-world existents not to have their well-being lowered from $L$ to $L^*$?[81]

What emerges here is that "solutions" to the non-identity problem and the problem of variation in population size are linked via the person-affecting principle. If one concludes that an outcome in which some individual does not exist *can* be worse for her—and thus that the person-affecting principle can be retained as a general principle for ranking outcomes, even where the population is not fixed—the "repugnant" conclusion will seem, on reflection, less repugnant. By shifting from $x$ to $y$ we achieve an outcome which would be better for many individuals—better than the zero level of nonexistence—at the cost of some well-off individuals being worse off. By contrast, from an impersonal perspective, it may seem that a world in which many people have lives barely worth living is ugly and squalid and always yields less value than a world in which a smaller population lives well, regardless of how much larger the population in the first case.

### *C. Infinite Population*

Imagine that each outcome in the outcome set contains an infinite number of individuals. This could arise in combination with a non-identity problem. But it could also arise quite independent of a non-identity problem. For example, imagine that an infinite number of individuals are born and exist in $x$, and the very same individuals are born and exist in $y$. Some of the individuals have different utility

---

80 However, it should be noted that difficulties arise in marrying "total" prioritarianism with the Atkinsonian SWF, $\frac{1}{1-\gamma}\sum_{i=1}^{N} u_i(x)^{1-\gamma}$, specifically in cases where the inequality aversion parameter $\gamma$ is greater than or equal to 1. In such cases, $g(0)$ is not zero, but rather is undefined. I will not attempt to resolve how the proponent of the Atkinsonian SWF (which is an attractive version of prioritarianism in fixed-population cases, for reasons I have discussed elsewhere, *see* sources cited *supra* note 1), might address this difficulty.

81 If one accepts the argument in this case, then it also surely works where the individuals in $x$ do not exist in $y$ and others take their place.

levels in $x$ than in $y$. How shall we rank the two outcomes?

A quite substantial body of scholarship in economic theory examines this sort of problem. This literature was initiated with pioneering articles by Frank Ramsey, Tjalling Koopmans, and Peter Diamond,[82] and focuses in particular on the problem of ranking countably infinite utility *streams* that extend forward in time from an initial period.[83]

---

82 *See* Peter A. Diamond, *The Evaluation of Infinite Utility Streams*, 33 ECONOMETRICA 170 (1965); Tjalling C. Koopmans, *Stationary Ordinal Utility and Impatience*, 28 ECONOMETRICA 287 (1960); F.P. Ramsey, *A Mathematical Theory of Saving*, 38 ECON. J. 543 (1928).

83 The economic theory literature on infinite utility streams is very large. Recent contributions include: Geir Asheim, Tapan Mitra & Bertil Tungodden, *A New Equity Condition for Infinite Utility Streams and the Possibility of Being Paretian*, *in* INTERGENERATIONAL EQUITY AND SUSTAINABILITY 55 (John Roemer & Kotaro Suzumura eds., 2007); Geir B. Asheim & Bertil Tungodden, *Resolving Distributional Conflicts Between Generations*, 24 ECON. THEORY 221 (2004); Claude d'Aspremont, *Formal Welfarism and Intergenerational Equity*, *in* INTERGENERATIONAL EQUITY AND SUSTAINABILITY 113 (John Roemer & Kotaro Suzumura eds., 2007); Kuntal Banerjee & Tapan Mitra, *On the Continuity of Ethical Social Welfare Orders on Infinite Utility Streams*, 30 SOC. CHOICE & WELFARE 1 (2008); Kuntal Banerjee, *On the Equity-Efficiency Trade off in Aggregating Infinite Utility Streams*, 93 ECON. LETTERS 63 (2006); Kuntal Banerjee, *On the Extension of the Utilitarian and Suppes-Sen Social Welfare Relations to Infinite Utility Streams*, 27 SOC. CHOICE & WELFARE 327 (2006) [hereinafter Banerjee, *On the Extension of the Utilitarian*]; Kuntal Banerjee & Tapan Mitra, *On the Impatience Implications of Paretian Social Welfare Functions*, 43 J. MATH. ECON. 236 (2007); Kaushik Basu & Tapan Mitra, *Aggregating Infinite Utility Streams with Intergenerational Equity: The Impossibility of Being Paretian*, 71 ECONOMETRICA 1557 (2003) [hereinafter Basu & Mitra, *Aggregating Infinite Utility Streams*]; Kaushik Basu & Tapan Mitra, *Possibility Theorems for Equitably Aggregating Infinite Utility Streams*, *in* INTERGENERATIONAL EQUITY AND SUSTAINABILITY 69 (John Roemer & Kotaro Suzumura eds., 2007); Kaushik Basu & Tapan Mitra, *Utilitarianism for Infinite Utility Streams: A New Welfare Criterion and its Axiomatic Characterization*, 133 J. ECON. THEORY 350 (2007) [hereinafter Basu & Mitra, *Utilitarianism for Infinite Utility Streams*]; Walter Bossert et al., *Ordering Infinite Utility Streams*, 135 J. ECON. THEORY 579 (2007); Graciela Chichilnisky, *An Axiomatic Approach to Sustainable Development*, 13 SOC. CHOICE & WELFARE 231 (1996); Juan Alfonso Crespo et al., *On the Impossibility of Representing Infinite Utility Streams*, 40 ECON. THEORY 47 (2009); Marc Fleurbaey & Philippe Michel, *Intertemporal Equity and the Extension of the Ramsey Criterion*, 39 J. MATH. ECON. 777 (2003); Chiaki Hara et al., *Continuity and Egalitarianism in the Evaluation of Infinite Utility Streams*, 31 SOC. CHOICE & WELFARE 179 (2008); Luc Lauwers, *Intertemporal Objective Functions: Strong Pareto Versus Anonymity*, 35 MATH. SOC. SCI. 37 (1998); Tapan Mitra & Kaushik Basu, *On the Existence of Paretian Social Welfare Quasi-Orderings for Infinite Utility Streams with Extended Anonymity*, *in* INTERGENERATIONAL EQUITY AND SUSTAINABILITY 85 (John Roemer & Kotaro Suzumura eds., 2007); Lars-Gunnar Svensson, *Equity Among Generations*, 48 ECONOMETRICA 1251 (1980); Yongsheng Xu, *Pareto Principle and Intergenerational Equity: Immediate Impatience, Universal Indifference, and Impossibility*, *in* INTERGENERATIONAL EQUITY AND SUSTAINABILITY 100 (John Roemer & Kotaro Suzumura eds., 2007); William Zame, *Can Intergenerational Equity be Operationalized?*, 2 THEORETICAL ECON. 187 (2007); Geir B. Asheim, Claude d'Aspremont & Kuntal Banerjee, *Generalized Time-Invariant Overtaking*, (Départment des sciences économiques de l'Université catholique de Louvain, Working Paper No. 2008039, 2008); Chiaki Hara, Kotaro Suzumura & Tomoichi Shinotsuka, *On the Possibility of Continu-*

Each stream $s$ has the structure $(u_1, u_2, \ldots)$, where $u_1$ is the utility in the first time period, $u_2$ in the second, and so forth. This literature is quite technical, but its essence is clear: it may well be impossible to rank infinite utility streams in a manner which is neutral between earlier and later utilities once we couple a time-neutrality requirement with other, seemingly plausible, axioms.

Time-neutrality can be formally expressed via an "anonymity" axiom, which says: if stream $s$ and stream $s^*$ are permutations of each other, then $s$ and $s^*$ must be ranked as equally good. In the infinity context, it is important to distinguish between different variants of the anonymity axiom. A weaker anonymity axiom says that $s$ and $s^*$ must be ranked the same if they are *finite* permutations of each other. A stronger anonymity axiom requires indifference to all permutations, including infinite permutations.[84]

Diamond establishes the following result.[85] Imagine that we aim to have a complete ordering of infinite utility streams. We require, further, that this ordering satisfy the principle of Pareto superiority, meaning in this context that if the utility in every period in stream $s^*$ is at least as large as the utility in stream $s$ in that period, and if the utility in at least one period in stream $s^*$ is strictly greater than the

---

*ous, Paretian, and Egalitarian Evaluation of Infinite Utility Streams* (Andrew Young Sch. of Policy Studies Research Paper Series No. 07-12, 2007), *available at* http://ssrn.com/abstract=976021.

There is also a smaller philosophical literature of infinity problems. This literature is more general than the economic theory literature just cited, because it discusses the problem of ranking worlds with an infinite number of "locations" of utility—whether or not those are organized in the form of a "stream" that begins with an initial period and never ends. *See, e.g.*, Donniell Fishkind et al., *New Inconsistencies in Infinite Utilitarianism: Is Every World Good, Bad or Neutral?*, 80 AUSTRALASIAN J. PHIL. 178 (2002); Luc Lauwers & Peter Vallentyne, *Infinite Utilitarianism: More Is Always Better*, 20 ECON. & PHIL. 307 (2004); Tim Mulgan, *Transcending the Infinite Utility Debate*, 80 AUSTRALASIAN J. PHIL. 164 (2002); Peter Vallentyne & Shelly Kagan, *Infinite Value and Finitely Additive Value Theory*, 94 J. PHIL. 5 (1997).

84 Formally, the weak anonymity axiom can be articulated as follows. Consider the set $T = \{1, 2, \ldots\}$, which denotes the periods in a utility stream. Function $p$ is a permutation if it is a bijection (one-to-one, onto mapping) from $T$ onto itself. It is a finite permutation if all but a finite number of elements of $T$ are mapped onto themselves. We can then say that stream $s^*$ is a permutation of stream $s$ if $u_t^{s^*} = u_{p(t)}^{s}$ for all $t$ and that stream $s^*$ is a finite permutation of $s$ if $p$ is a finite permutation. (The term "$u_t^s$" simply means the utility in stream $s$ in period $t$.) The weak anonymity axiom requires that $s^*$ and $s$ be ranked as equally good whenever $s^*$ is a finite permutation of $s$. *See, e.g.*, Asheim & Tungodden, *supra* note 83, at 223; Bossert et al., *supra* note 83, at 581; Fleurbaey & Michel, *supra* note 83, at 782; Svensson, *supra* note 83, at 1252. The strong anonymity axiom requires indifference to any permutation, even one that is not finite. *See, e.g.*, Fleurbaey & Michel, *supra* note 83, at 782–83.

For a full discussion of possible anonymity requirements in the infinity context, and their consistency with various forms of the Pareto principle, see *id.*

85 *See* Diamond, *supra* note 82.

utility in $s$ in that period, then $s^*$ must be ranked higher than $s$.[86] If we further stipulate that this ordering satisfy a *continuity* property as well as weak anonymity, impossibility results. No ordering of infinite streams can be complete, continuous, satisfy Pareto superiority, and satisfy even the weak anonymity axiom of indifference to finite permutations. Building on this striking finding, Basu and Mitra show that there is no ordering of utility streams which is complete, representable by a SWF, satisfies Pareto superiority, and also satisfies the weak anonymity axiom.[87]

Do these findings wholly undermine the desire to preserve time-neutrality once we enter the domain of infinite utility streams? The Diamond and Basu/Mitra articles focus on the problem of producing a *complete* ordering of infinite streams; this is true of much, if not all, of the literature. But it is far from clear why completeness is a normatively compelling axiom.[88] As I suggested earlier, and have argued at length elsewhere,[89] even in the case of a fixed, finite population, an attractive moral theory such as prioritarianism may produce only an incomplete ordering of outcomes—a so-called "quasi-ordering."

Once we relinquish the completeness requirement, and require only that the ranking of infinite utility streams be a *quasi-ordering*, the ambition to reconcile a measure of time-neutrality and the Pareto principle becomes more feasible. Svensson notes that there are a variety of criteria for quasi-ordering streams that are consistent both with the weak anonymity axiom and with the Pareto superiority principle.[90] One such criterion is the so-called "overtaking" criterion. The idea is that one stream "overtakes" another if, after a certain number of periods, the sum of utility from the beginning is always larger with the first stream. More formally, the overtaking criterion says: stream $s$ is at least as good as stream $s^*$ iff there exists a time $T^*$ such that, for all $T > T^*$, $\sum_{t=1}^{T} u_t^s \geq \sum_{t=1}^{T} u_t^{s^*}$, where $u_t^s$ is the utility in

86 Throughout the discussion of infinite utility streams, this is what I mean by the principle of Pareto superiority.

87 *See* Basu & Mitra, *Aggregating Infinite Utility Streams*, *supra* note 83.

88 Some scholars in the infinite-utility-stream literature have investigated the possibility of relaxing completeness. *See, e.g.*, Asheim & Tungodden, *supra* note 83; Basu & Mitra, *Utilitarianism for Infinite Utility Streams*, *supra* note 83; Bossert et al., *supra* note 83; Hara et al., *supra* note 83.

89 *See* ADLER, WELL-BEING AND EQUITY, *supra* note 1; *see also* ADLER & POSNER, *supra* note 18, at 161–62 (arguing that outcomes may be incomparable with respect to overall well-being).

90 *See* Svensson, *supra* note 83; *see also* Bossert et al., *supra* note 83, at 579–80.

stream $s$ in period $t$.[91]

Note that there is no discount factor in this formula; it looks, rather, to the undiscounted sum of utilities. If, instead, we compare utility streams using the simple discounted sum of utilities $\sum_{t=1}^{\infty} D(t)(u_t^s)$, there is a violation of the weak anonymity axiom.

Geoffrey Heal has criticized the overtaking criterion because it fails to be neutral between the following sort of pair of utility streams. Imagine that $s = (1, 0, 1, 0, \ldots)$ and $s^* = (0, 1, 0, 1, \ldots)$. Then $s$ is ranked as better than $s^*$ by the overtaking criterion but time-neutrality would seem to require that $s$ and $s^*$ be ranked as equal.[92]

Note, however, that indifference between $(1, 0, 1, 0, \ldots)$ and $(0, 1, 0, 1, \ldots)$ is not required by the weak anonymity axiom. The two streams are not finite permutations of each other; they are *infinite* permutations. Neutrality between $(1, 0, 1, 0, \ldots)$ and $(0, 1, 0, 1, \ldots)$ is required by the *strong* anonymity axiom. However, it can be shown that there is no way to produce even a quasi-ordering of utility streams which (1) respects the principle of Pareto superiority and (2) respects the strong anonymity axiom, i.e., indifference to infinite permutations. The following example illustrates that strong anonym-

---

91 The overtaking criterion originates with von Weizsacker & Atsumi. *See* Hiroshi Atsumi, *Neoclassical Growth and the Efficient Program of Capital Accumulation*, 32 REV. ECON. STUD. 127 (1965); Carl Christian von Weizsacker, *Existence of Optimal Programs of Accumulation for an Infinite Time Horizon*, 32 REV. ECON. STUD. 85 (1965). For recent discussions, see Asheim & Tungodden, *supra* note 83, at 226–28; Banerjee, *On the Extension of the Utilitarian*, *supra* note 83, at 333–35; Basu & Mitra, *Utilitarianism for Infinite Utility Streams*, *supra* note 83, at 356–63; Fleurbaey & Michel, *supra* note 83, at 786–87; Svensson, *supra* note 83, at 1253. For a variation on the overtaking criterion, see Asheim et al., *supra* note 83 ("generalized time-invariant overtaking").

92 *See* GEOFFREY HEAL, VALUING THE FUTURE: ECONOMIC THEORY AND SUSTAINABILITY 65–67 (1998). Note that, for each odd value of $T$, $\sum_{t=1}^{T} u_t^s > \sum_{t=1}^{T} u_t^{s*}$ —where $s = (1, 0, 1, 0, \ldots)$ and $s^* = (0, 1, 0, 1, \ldots)$. For every case where $T$ is even, $\sum_{t=1}^{T} u_t^s = \sum_{t=1}^{T} u_t^{s*}$. Therefore, if we set $T^*$ equal to zero, for all $T > T^*$, the sum of utilities up to period $T$ is at least as large (greater than or equal) in $s$ as in $s^*$. However, it can be seen that there is *no* $T+$ such that, for all $T > T+$, $\sum_{t=1}^{T} u_t^{s*} \geq \sum_{t=1}^{T} u_t^s$ —because the cumulative sum of utilities in $s^*$ is always less than that in $s$ in odd periods.

Because $s$ is at least as good as $s^*$ by the overtaking criterion, and $s^*$ is not at least as good as $s$, it follows that $s$ is better than $s^*$ by the overtaking criterion—which is what Heal in turn criticizes.

ity and the principle of Pareto superiority are mutually inconsistent.[93]

**Table 2. The Conflict Between Anonymity and Pareto Superiority**

| | Period | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | . . . |
| **Stream *s*** | 1 | 3/2 | 1/3 | 7/4 | 1/5 | 11/6 | 1/7 | 15/8 | 1/9 | . . . |
| **Stream *s**** | 1 | 7/4 | 3/2 | 11/6 | 1/3 | 15/8 | 1/5 | 19/10 | 1/7 | . . . |

Strong anonymity requires that the two streams be ranked as equally good, while the Pareto superiority principle requires that the second stream be ranked as better than the first.

In short, criticisms of the overtaking criterion that point to its lack of indifference in cases of infinite permutations—such as Heal's—are not particularly persuasive, because what these cases really evidence is a generic impossibility, rather than a specific flaw of the overtaking criterion.

A different point is that the overtaking criterion is a *utilitarian* criterion. It compares two streams by comparing the sum of utilities up to each point in time. However, it is straightforward to produce a *prioritarian* variant of the overtaking criterion which sums a strictly increasing and strictly concave *g*-function of utilities. Prioritarian overtaking says that stream *s* is at least as good as stream *s** iff there is a time *T** such that, for all $T > T^*$, $\sum_{t=1}^{T} g(u_t^s) \geq \sum_{t=1}^{T} g(u_t^{s*})$.

To be sure, the prioritarian criterion thus formulated is applicable to the problem of ranking infinite utility streams. For purposes of social choice, we need to rank outcomes. In the infinite-stream setup, there is a single utility in each period. But a framework for intertemporal social choice should, of course, allow for the possibility that *multiple* individuals, possibly at different levels of (lifetime) utility, may exist at the same time.

---

93 Stream *s* is produced by taking the alternating sequence 0, 1/2, -2/3, 3/4, -4/5, 5/6, . . . , and adding 1 to each term. Stream *s** is produced by rearranging the terms of stream *s* as follows: keep the first term in place, shift the second term (3/2) one to the right, and, for every other term shift it two to the right if its denominator is odd, and two to the left if its denominator is even.

On the inconsistency between Pareto superiority and strong anonymity (indifference to infinite permutations), see Asheim & Tungodden, *supra* note 83, at 229. Asheim and Tungodden provide a simpler example of the conflict. The example I provide here, however, involves a case in which one stream is a permutation of another, yet is strictly better in all but a single time period.

However, it is easy to generalize the prioritarian overtaking criterion to cover this case. Assume that the temporal structure of the universe is the same in all outcomes: time begins with an initial period, and continues *ad infinitum*. In each outcome in each period, there are a finite number of individuals who are born, possibly zero; this number need not be the same in all the outcomes.[94] Let us say that $N(x, t)$ is the number of individuals born in outcome $x$ during or before period $t$. In a given outcome, assign the individuals who exist in that outcome numbers corresponding to their date of birth. So individuals numbered 1 to $N(x,1)$ in outcome $x$ are born in period 1 in outcome $x$; individuals numbered $N(x,1) + 1$ to $N(x,2)$ are born in period 2, and so forth. Then the prioritarian overtaking criterion says that $x$ is at least as good as $y$ iff there is a time $T^*$ such that, for all $T > T^*$, $\sum_{i=1}^{N(x,T)} g(u_i(x)) \geq \sum_{i=1}^{N(y,T)} g(u_i(y))$. As in the core case of a fixed and finite population, $u_i(x)$ is the lifetime utility of individual $i$ in outcome $x$.

In the core case of a fixed and finite population, as noted earlier, the prioritarian criterion $\sum_{i=1}^{N} g(u_i(x))$ satisfies a number of axioms: the Pareto indifference principle, the principle of Pareto superiority, the Pigou-Dalton principle, separability across individuals, and anonymity.[95] In the new setup, involving an infinite future, the prioritarian overtaking criterion also satisfies Pareto indifference, Pareto superiority, Pigou-Dalton, and separability across individuals.[96] As for ano-

---

94 However, I will assume that an individual is born on the very same date in all outcomes where she exists. Although this assumption is not entailed by the nature of personal identity, *see supra* text accompanying note 44, relaxing the assumption creates difficulties for the prioritarian overtaking criterion as presented here. Note that, if some individual is born later in outcome $x$ than outcome $y$, outcome $x$ might be Pareto superior to $y$ and yet the sum $\sum_{i=1}^{N(x,T)} g(u_i(x))$ might be less than the sum $\sum_{i=1}^{N(y,T)} g(u_i(y))$, for values of $T$ in between the individual's birth date in $y$ and $x$. This might yield inconsistencies between the prioritarian overtaking criterion and the principle of Pareto superiority if an infinite number of individuals have different birth dates in $x$ and $y$. Similar difficulties can arise with the Pareto indifference principle. How it might be possible to relax the assumption that an individual is born on the very same date in all outcomes where she exists is a question I leave for another day.

95 *See supra* Part I.

96 For purposes of applying these criteria to an outcome set with potential nonexistents, who exist in some but not all outcomes, such individuals should be assigned a utility of zero in the outcomes where they do not exist. Then Pareto indifference requires that outcome $x$ be ranked as equally good as $y$ if (1) every individual who exists in both outcomes is equally well off in both, and (2) every individual who exists in one but not both outcomes has a utility of zero in the outcome where she exists. Pareto superiority requires that outcome $x$ be ranked as better than outcome $y$ if the utility assigned each individual in $x$ is at least as great as the utility assigned her in $y$, and there are some individuals assigned strictly greater utility—where, by "the

nymity, it fails to be indifferent to infinite permutations of utility, but it is indifferent to finite permutations.[97] On balance, this seems a plausible approach for prioritarians to use in ranking outcomes involving an infinite future.

To be sure, the prioritarian overtaking criterion is not a complete solution to infinity problems. It tells us how to rank an outcome set in which time has a beginning but no end in all outcomes. It gives no guidance in the following sorts of cases: (1) time has a beginning and continues *ad infinitum* in some outcomes, while in others it has a beginning and ends after a finite number of periods (this is a plausible outcome set if the decisionmaker is *uncertain* whether the universe will end); (2) time has no beginning and no end in some outcomes; (3) the population is *infinite* in some periods in some outcomes (this may perhaps be a plausible outcome set if the decisionmaker is interested, not just in the well-being of the persons who live on Earth—which has a finite carrying capacity—but the well-being of all intelligent life in the universe).

A different possible reaction to infinity problems is to deny their practical relevance. For example, if we are constructing a framework intended to provide guidance in impartially considering the interests of the human population of the Earth, and if we can be reasonably certain not only that Earth had a beginning but also that Earth will end, then we can be reasonably certain that the morally relevant population in every outcome will be finite. Of course, it might be objected that this construal of the morally relevant population is too restricted. If individuals born on Earth migrate to Mars, shouldn't we

---

utility assigned to each individual," I mean that individual's utility if she exists, or zero if she does not. Pigou-Dalton means that if $x$ and $y$ are identical, except that $u_i(y) = u_i(x)+\Delta u$, $u_j(y) = u_j(x)-\Delta u$, $u_j(x) > u_j(y) > u_i(y) > u_i(x)$, where these utilities are the utilities assigned to individuals $i$ and $j$ in outcomes $x$ and $y$, then $y$ is better than $x$. Separability means that the ranking of $x$ and $y$ does not depend on which utilities are assigned to individuals who are assigned the same utilities in both outcomes.

97 By these terms, in this context, I mean the following. Given an outcome set, there is a set of all individuals who exist in at least one of the outcomes. A permutation is a bijection (one-to-one, onto mapping) from this set onto itself. A finite permutation is a permutation with the property that only a finite number of individuals are mapped onto different individuals. An infinite permutation is a permutation which is not finite.

The utilities assigned individuals in outcome $y$ are a permutation of the utilities assigned individuals in outcome $x$ if there is some permutation $p$ such that each individual $i$'s utility in $x$ is the utility of individual $p(i)$ in $y$. If there exists a finite such $p$, then $y$ is a finite permutation of $x$. Less formally, $y$ is a finite permutation of $x$ if there is some finite subset of all the individuals (those who exist in at least one outcome), such that the utilities assigned to those individuals in $y$ are a rearrangement of the utilities assigned to them in $x$, and all individuals not in the subset are assigned the very same utilities in $y$ and $x$.

care about them? But we might be reasonably certain that this universe will end, and thus that the set of individuals born on Earth, plus their descendants, will be finite in any reasonably possible outcome.

### *Conclusion*

This Article has considered the problem of future generations within the context of prioritarianism: a moral view which is welfarist but sensitive to equity. Unlike utilitarianism, a prioritarian approach gives greater weight to well-being changes affecting worse-off individuals.[98] On my specific rendering, prioritarianism sees worse-off individuals as having stronger *claims* to well-being than better-off individuals.

The Article addressed both the "core case," in which the world's intertemporal population is fixed and finite[99] (the same $N$ individuals exist in all possible outcomes, with $N$ a finite number), and departures from the "core case" discussed in the philosophical and economic literature, namely non-identity cases,[100] variation in the size of the intertemporal population,[101] and an infinite future.[102]

I argued for neutrality between generations in the "core case." In that case, outcomes should be ranked using the formula $\sum_{i=1}^{N} g(u_i(x))$, rather than a formula that discounts the well-being of future generations, such as $\sum_{i=1}^{N} D(b(x;i))g(u_i(x))$, with $D(.)$ a discount factor that depends on the date individuals are born. A neutralist approach that ranks policies using the basic formula without a discount factor, $\sum_{i=1}^{N} g(u_i(x))$, is *already* sensitive to various important considerations raised in the literature on discounting, namely opportunity costs, uncertainty, the fact that future generations are likely to be richer, and worries about impoverishing the present for the sake of the future.[103] Reciprocally, discounting the utilities of future generations would arguably violate the Pareto indifference axiom and, in any event, would violate the anonymity axiom.[104] Anonymity is a formal, welfarist, ex-

98 *See supra* Part I.

99 *See supra* Part II.

100 *See supra* Part III.A.

101 *See supra* Part III.B.

102 *See supra* Part III.C.

103 *See supra* text accompanying notes 46–56.

104 *See supra* text accompanying notes 42–45.

pression of the idea that moral reasoning should be impartial between persons. It says that if outcome $x$ produces the very same pattern of well-being as outcome $y$ (i.e., the utility vector for $x$ is a permutation of the utility vector for $y$), then $x$ and $y$ are equally good.

Matters become more complicated when we move outside the "core case." The Article does not take a firm position on how prioritarianism should handle non-identity cases, variation in population size, and an infinite future. I did, however, tentatively suggest that neutrality between current and future generations can plausibly be preserved even in such cases: in the first two instances, via the "total" prioritarian formula $\sum_{i=1}^{N(x)} g(u_i(x))$ (notwithstanding the "repugnant conclusion");[105] and in the last instance via a prioritarian version of the "overtaking" criterion, which is at least consistent with a weak version of the anonymity axiom.[106]

105 *See supra* text accompanying notes 64–67, 80–81.

106 *See supra* text accompanying notes 96–97.