Should We Discount the Welfare of Future Generations? Ramsey and Suppes versus Koopmans and Arrow

Graciela Chichilnisky, Peter J. Hammond & Nicholas Stern

(This paper also appears as CRETA Discussion paper 43

& also CAGE Discussion paper 386)

August 2018 No: 1174

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

Should We Discount the Welfare of Future Generations?
Ramsey and Suppes versus Koopmans and Arrow

Graciela Chichilnisky: gc9@columbia.edu
Dept. of Economics, International Affairs Building, Columbia University,
420 W 118 St MC 3308, New York, NY 10027, USA

Peter J. Hammond: p.j.hammond@warwick.ac.uk
Dept. of Economics, and CAGE (Competitive Advantage
in the Global Economy), University of Warwick, Coventry CV4 7AL, UK

Nicholas Stern: n.stern@lse.ac.uk
Dept. of Economics, and Grantham Research Institute on Climate Change
and the Environment, LSE, Houghton Street, London WC2A 2AE, UK

Typeset from chicHamSternArrow9c.tex dated 2018 August 14th.

Abstract: Ramsey famously pronounced that discounting "future enjoyments" would be ethically indefensible. Suppes enunciated an equity criterion implying that all individuals' welfare should be treated equally. By contrast, Arrow (1999a, b) accepted, perhaps rather reluctantly, the logical force of Koopmans' argument that no satisfactory preference ordering on a sufficiently unrestricted domain of infinite utility streams satisfies equal treatment. In this paper, we first derive an equitable utilitarian objective based on a version of the Vickrey–Harsanyi original position, extended to allow a variable and uncertain population with no finite bound. Following the work of Chichilnisky and others on sustainability, slightly weakening the conditions of Koopmans and co-authors allows intergenerational equity to be satisfied. In fact, assuming that the expected total number of individuals who ever live is finite, and that each individual's utility is bounded both above and below, there is a coherent equitable objective based on expected total utility. Moreover, it implies the "extinction discounting rule" advocated by, *inter alia*, the *Stern Review* on climate change.

Keywords: Discount rate, utilitarianism, consequentialization, Vickrey–Harsanyi original position, Suppes criterion, optimal population, average versus total utility, intergenerational equity, long-run discounting, sustainable preferences, extinction discounting rule.

JEL Classification: D63, D70, D90, Q54, Q56

1 Introduction

1.1 Kenneth Arrow on Discounting the Future

On February 23rd 2017 the world lost Kenneth Arrow (KJA), the father of social choice and a towering giant among economists. He was greatly admired by all three authors of this paper.[1]

During his seventies KJA had given two closely related lectures, first to the 1995 World Congress of the International Economic Association, and second to a 1996 conference on discounting that was organized jointly by the (Stanford) Energy Modeling Forum and by Resources for the Future. These lectures, later published in Arrow (1999a, b), both considered the question we address in this paper: at what rate should we discount the welfare of future generations? Indeed, should the discount rate even be zero?

In these two lectures, KJA highlighted two arguments in favour of discounting future generations:[2]

1. a *strong* argument, claiming that failing to discount future generations would lead to a logical inconsistency;

2. a *weak* argument, claiming that failing to discount future generations would imply a savings rate that imposed excessive sacrifices on the current generation.

Our main concern in this paper is whether, in a suitably extended utilitarian social choice theory, the strong argument can be obviated by recognizing that there is a stochastic process, perhaps partly subject to human influence, which determines when humanity will eventually become extinct. In fact we specify two assumptions ensuring that maximizing the expected value of the unweighted total utility of all future generations is logically coherent as a social welfare objective. These assumptions are:

[1] Gaertner (2017), a founding editor of *Social Choice and Welfare*, contributed a short obituary to these pages. It was preceded by a photograph that one of us (PJH) had taken more than 10 years earlier of KJA standing outside the Landau Economics Building at Stanford. It portrayed him holding a representation in metal of an Edgeworth box with an apple core inside, bearing the label "The Core of the Arrow–Debreu Economy". This is a sports trophy awarded each year to the winners of the "little big game" of (American) touch football that is played between the Ph.D. students of the Departments of Economics at Berkeley and Stanford.

[2] KJA attributed both these arguments to Koopmans. Yet later, in Section 2.5, we will give reasons for believing that Koopmans, as well as his co-author Diamond, thought that these arguments should be applied only to the intertemporal preferences of a single consumer, rather than to those of an "ethical benefactor" who is choosing policy on behalf of society as a whole.

1. the expected total population of all individuals who live at any time before the uncertain extinction date is finite;
2. and, following an assumption made famous by Arrow (1951, 1965, 1971, 1972), the common fundamental utility function of each potential individual is bounded both above and below.

With these assumptions, maximizing expected total utility involves the *extinction discounting rule* whereby successive future generations are weighted by their declining probability of existence. Under this rule, which is advocated *inter alia* in the *Stern Review* (Stern, 2006), all individuals are treated equally, contingent on their existence, as intergenerational equity would seem to require. In fact, future generations should be given increasing weight, in proportion to their growing size, as long as expected population, allowing for the probability of extinction, continues to expand.

Finally, we consider a refinement of the extinction discounting rule which meets Chichilnisky's (1996, 1997, 2009) definition of "sustainable preferences".

1.2 Background

Despite the arguments set out in Arrow (1999a, b), as well as the discussion in Arrow (2007) of the *Stern Review* (Stern, 2006), it seems that KJA's views on discounting the welfare of future generations may have been evolving. During the late 1960s, he had begun working on the problem of specifying an appropriate discount rate in social cost–benefit analysis. This led to the eventual publication of Arrow (1966, 1982), as well as of Arrow and Kurz (1970) and Arrow and Lind (1970).[3]

Some decades later, while director of the Stanford Institute of Theoretical Economics (SITE), he organized in 1993 a summer workshop session on the topic "Reconsideration of Values".[4] It included GC's invited talk, which PJH attended, with an early version of ideas that later appeared in

[3]NHS recalls that he first met KJA in Oxford in 1969, on a visit organized by Mirrlees. This was around the time when Little and Mirrlees (1969) was published. There was an intense discussion of values, intertemporal and otherwise, in the context of social cost–benefit analysis. KJA was, as ever, in the vanguard.

[4]PJH recalls that in 1988, as KJA was nearing what at the time was the compulsory retirement age of 70, while John Shoven was chair of the Stanford Department of Economics, both Donald Brown and Paul Milgrom were recruited. Also IMSSS (the Institute of Mathematical Studies in the Social Sciences), which Mordecai Kurz had directed very successfully for 20 years, was replaced by SITE, with Milgrom as its founding director. SITE remains very active 30 years later.

Chichilnisky (1996, 1997).[5] Rather than follow earlier work described in the early part of Section 7 which had considered whether infinite-horizon plans are sustainable, she instead characterized “sustainable preferences”. These were the welfare criteria defined in Section 7 which avoid dictatorship of either the present or the future, and whose maximization would normally lead to some form of sustainable development path.

1.3 An Unfulfilled Plan

On PJH’s frequent return visits to Stanford, during several discussions over lunch, KJA put forward two reasons to discount the welfare of future generations. Essentially these correspond to two arguments that KJA attributed to Koopmans, as discussed further in Section 2.[6]

Indeed, following one of these lunches, KJA seemed ready to go along with a suggested plan for PJH to write up, as the genesis of a joint paper, an argument for what we call here “extinction discounting”, as discussed in Section 1.1. On the basis of arguments from social choice theory, this is the rule which we advocate in this paper, subject to a possible caveat regarding GC’s concept of “dictatorship of the present”, as discussed in Section 8.7.

The intention had been that, in a fuller version of this projected joint work, eventually KJA would respond with any objections he might have to the extinction discounting rule. Perhaps KJA would also have taken up his side of some earlier friendly debates with NHS, while also engaging with GC’s work on issues surrounding sustainable preferences that KJA had cited. It was also planned that this two-author paper would become a tribute, perhaps even in these pages, to our mutual friend Patrick Suppes, the great scientific philosopher, who had died on November 17th, 2014.[7] Indeed, the equity principle due to Suppes plays a key role in our argument.

Unfortunately PJH’s other commitments prevented this plan from being taken even as far as a first draft. What can never be determined, therefore, is whether KJA’s conversations with PJH just possibly might have been a

[5]See also the first footnote of Chichilnisky (1996). A preprint version of Chichilnisky (1997) was cited in Arrow, Cline *et al.* (1996).

[6]During these discussions PJH remained unaware of Arrow (1999a, 1999b), whose existence KJA did not deign to mention. But it was clear that KJA had been inspired by his late friend Tjalling Koopmans, who had been a mentor during his time at the Cowles Foundation when it was still in Chicago. It is also said that Koopmans had persuaded KJA and Gérard Debreu to amalgamate their separate working papers into what became Arrow and Debreu’s (1954) classic paper on existence of general competitive equilibrium.

[7]We should mention that, in addition to KJA, Suppes had also been an advisory editor of *Social Choice and Welfare*.

typical sign that he recognizing the importance of revisiting the topic of discounting in order to make sure that both the ethical and mathematical analysis were done right. Indeed, by combining a simplistic misuse of the strong argument regarding logical necessity with claims that market interest rates carry relevant information, many writers had tried to justify excessively high rates of pure-time discounting. Our paper deals directly with the strong argument. As for market interest rates, they do not clearly reflect social preferences, especially as most financial markets suffer from important market failures.[8]

1.4 Outline of Paper

Section 2 sets out in some detail Arrow's (1999a, b) discussion of the two arguments in favour of discounting the welfare of future generations that he attributed to Koopmans. Of these, the one that deserves more attention from social choice theorists is the first "strong" argument claiming that discounting is logically necessary. It will therefore be the main subject of this paper. Limitations of space and time compel us to offer no more than a few remarks in Section 2.6 regarding the second "weak" argument stating that failure to discount asks the current generation to sacrifice too much. See, however, Stern (2006, 2008, 2015) for extensive further discussion.

Next, Section 3 begins to lay out the basic normative framework that we use to analyse the ethical decisions involved in choosing between alternative futures for humanity. We emphasize that our approach is thoroughly normative, and so departs from any descriptive framework that involves concepts such as Hume's (1739) distinction between higher and lower selves, or Kahneman's (2011) distinction between fast and slow thinking. Hume in particular, and probably Kahneman as well, saw the importance of basing policy recommendations on social values. We have in mind a concept of rationality closely related to the kind of thoughtful behaviour discussed in the early pages of Savage (1954), after he realized that he had at first reported irrational preferences in the context of the Allais paradox.

Also, like most work in social choice and welfare, our approach is thoroughly consequentialist, for reasons that we try to explain in Section 3.2. Nevertheless, Section 9.2 offers some discussion of alternative ethical approaches, as well as the conclusion reached in Stern (2014a, 2015) that these approaches also do not justify the bias toward the present that discounting the welfare of future generations would imply.

[8]See Hammond (1992) for one possible explanation of these failures, based on a combination of moral hazard and adverse selection in loan markets.

Reverting to social choice theory, we note that Arrow (1950) and many subsequent works followed the tradition pioneered by Borda and Condorcet, which based social preferences entirely on individual preferences. To avoid the logical difficulties that this approach created, Sen (1970) pioneered the use of richer information in the form of interpersonal comparisons of utility. Yet, as discussed in Hammond (1996b), there was no clear explanation of how these comparisons should be made.

Here, by contrast, as befits an analysis of how (and whether) to discount the welfare of future generations, we revisit and elaborate some of the ideas concerning "interpersonally comparable utility" that were set out first in Hammond (1991) and then more systematically in Fleurbaey and Hammond (2004). Section 3 explains how, when the welfare of each individual depends upon their own all-inclusive personal lifetime consequence $y \in Y$, one can derive a single cardinal equivalence class of interpersonally comparable fundamental utility functions $Y \ni y \mapsto u(y) \in \mathbb{R}$. These are defined on the universal domain Y, with expectations that apply to the set $\Delta(Y)$ of all possible lotteries over that domain. Indeed, when choosing between members of a family of biased versions of the "Vickrey–Harsanyi original position", it is ethically appropriate to maximize the expectation of u.

Starting with this basic framework for our version of utilitarian social choice theory, Section 4 presents a version of Suppes' equity criterion. This implies that, given a fixed population, it is appropriate to consider an unbiased Vickrey–Harsanyi original position, where there is an equal probability of becoming any one of the individuals in society.

Thereafter, Section 5 argues that when the population is not fixed, it is appropriate to maximize the expected total welfare of all individuals who ever exist in the future. Unlike maximizing their expected average welfare, this is linked to the idea that, in a relevant initial position where even each individual's eventual existence is treated as risky, an increased population implies a higher probability of coming into existence.

Given this measure of welfare for each potential individual in society, it remains to determine what intergenerational objective is ethically appropriate for the actual social decisions that we will be forced to make. We note that these decisions must allow for the problem of climate change, including worries about the *Storms of My Grandchildren* that James Hansen has written about. This task is made much more complicated to the extent that the strong argument in favour of discounting the lives of future generations is valid.

Now, an Arrow social welfare function is a mapping from a domain of individual preference profiles into a social preference ordering. A key postulate used in Arrow's impossibility theorem is that the social welfare function has an unrestricted domain of all logically possible preference profiles. A well-known escape from this theorem involves restricting the domain of the Arrow social welfare function. Specifically, under the assumption that there is an odd finite number of individuals, if the domain of preference profiles is restricted to those that are single-peaked relative to a given linear ordering of the social states, then the Condorcet majority rule is a social welfare function that satisfies all four of Arrow's conditions. In the same way, Sections 6 and 7 together explore various restrictions on the domain of infinite horizon decision trees. Indeed, when the welfare measure for each generation is its total level of welfare, even discounting future generations' welfare only works if population growth is slow enough for the discounted total population to be finite.

We are especially interested, however, in attempts to escape what Chichilnisky (1996, 1997, 2009b) has called "dictatorship of the present", which is defined in Section 7.5. These attempts include: a Ramsey rule without discounting on a restricted domain; maximin on a restricted domain; various extensions of maximin due to Asheim, Mitra and Tungodden (2012), and Asheim and Zuber (2013); and the general family of "sustainable" preferences advocated by Chichilnisky (1996, 1997, 2009b) and Heal (1998).

Next, in Section 8, we look for circumstances in which after all one can justify, if not full intergenerational equity without discounting, then at least the "extinction discounting" rule that, after receiving early attention in Ramsey (1928) and Mirrlees (1967), is favoured in the *Stern Review* — see Stern (2006, 2015) and, among many critical assessments, Arrow (2007) and Dasgupta (2007, 2008). In our framework, which combines a stochastic extinction process with personal consequence streams for infinitely many potential future individuals, the key domain restriction is to require that the expected population, before ultimate extinction, is finite.

The important assumption of a finite expected population prevents the infinite tail of future generations from wagging the dog of current policy formation. It is obviously satisfied under the plausible hypothesis that total world population eventually stabilizes, provided that there is a positive probability of extinction at dates far enough into the future, even if that probability is arbitrarily small. Indeed, what our domain restriction requires is that the increasing probability of extinction should outweigh population growth. So, provided that the probability of extinction approaches 1 fast enough, this assumption does not exclude exponential population growth of the kind that one finds in some versions of the standard economic growth model due to Solow (1956) and Swan (1956) — growth which Ehrlich (1968) described as a "population bomb".

Section 9 sets out some tentative conclusions. In particular, we speculate whether KJA's eagerness to revisit the discounting issue may have reflected some doubt about what he had written in Arrow (1999a, b). We also briefly discuss discounting in the light of ethical doctrines other than utilitarianism. Finally, we offer some thoughts about how technical progress, including the kind of learning by doing that Arrow (1962) had popularized, might affect the "weak argument" that failure to discount future generations' welfare — or failure to discount it at rates higher than those implied by extinction discounting — might place an excessive burden on current generations.

2 Two Arguments for Discounting

2.1 The Social Rate of Discount

Following what has become standard practice since at least Arrow and Kurz (1970), Section 3 of Arrow (1999a) gives the formula $r = \rho + \theta g$ for the social rate of discount, where[9]

> ρ is the rate of pure time preference (if any), θ is the elasticity of marginal utility with respect to income, and g is the rate of growth of consumption *per capita* ... $\rho = 0$ implies equal treatment of present and future.

Then Section 4 of Arrow (1999a) starts as follows:

> In [the] formula [$r = \rho + \theta g$], the second term, θg, is, I think, fairly uncontroversial. If future individuals are going to be better off than we are, then our willingness to sacrifice on their behalf is certainly reduced. It would require a greater rate of return to justify our depriving ourselves of consumption.
>
> But the presence of pure time preference, denoted by ρ, has been very controversial. The English economists, in particular, have tended to be very scornful of pure time preference.

2.2 Discounting Future Consumption

In the ensuing discussion of the term θg, Arrow (1999a) considers how to value an increment in a good in the future, relative to an increment now.

[9]See also Stern (1977).

This is the discount factor, which we denote by β, for that good at that time. The proportionate rate of fall of the discount factor, given by $-\dot{\beta}/\beta = -\frac{\mathrm{d}}{\mathrm{d}t}\ln\beta$, is the discount rate for that good at that time; it clearly depends on both the good and the time. In our view, the focus in economic assessments should be on the discount factor, as that is the key shadow price, relative to now, which is needed to find the marginal present value of any change in costs or benefits occurring at any specific time in the future. When we need to evaluate a stream of costs and benefits over time, we can consider the net present value (NPV) of the whole stream, with the costs and benefits at each time t weighted by the discount factor $\beta(t)$. That is, at time 0 one considers $\text{NPV} := \int_0^T \beta(t)b(t)\,\mathrm{d}t$, where $b(t)$ denotes net benefit at time t, and T denotes the terminal time. When allowing for the inevitable uncertainty surrounding future costs and benefits, one approach is to consider their expected discounted value.

Once we have the right concept of the discount factor that should be applied to future consumption, it becomes immediately clear that this factor will depend on the state of affairs at each relevant time in the future. Unmanaged climate change could make future generations very poor. Then we might place a very high value on extra goods that are available in calamitous circumstances. This could even imply negative a discount rate, or equivalently, a discount factor greater than one. This possibility also makes it clear that each uncertain future state of the world that could occur is of critical relevance. So, too, is the person or persons who may experience increments in income. Indeed, using the term "*the* discount rate", as if there is just one given rate, clearly constitutes a serious misunderstanding of the basic issues.

So the relevant discount rates in any calculation of expected discounted value are endogenously determined as a result of our planned decisions. Moreover, this endogeneity is potentially severe in the case of climate change. We conclude that, unlike examples like border prices when considering policies in what international economics regards as a small national economy, one cannot simply import an exogenous discount rate from outside the model.

2.3 Discounting Other Generations' Future Lives

Our concern in this paper, however, is much more with the ρ term of the formula $r = \rho + \theta g$ that Arrow (1999a) gives for the discount rate. This is often called pure time discounting, or pure time preference. It arises when we contemplate policies whose effects, like the climate change induced by greenhouse gas emissions, extend far into the future. This should force us to give some value to the consumption of people who live in the future. Pure time discounting involves, and is even essentially defined as, the relative weight attached to a life in the future compared to a life now, when the two lives are otherwise identical in all respects. That is, the only difference is that one life is in the future, whereas the other is right now.

If the pure time discount rate were 2% per annum, for example, then a life starting 35 years in the future that is otherwise identical to a life that starts now, would have a relative value of $1.02^{-35} \approx 0.5$ compared to a life that starts now. In this sense we are "discounting future lives", which is effectively discriminating by date of birth. It cannot be justified by some notion of the future life being better because it has higher consumption; that would be discounting future consumption, as considered in Section 2.2, as opposed to discounting future welfare *per se*. We emphasize that we are making an ethical comparison between two lives that are identical, except for the dates of birth. It is very difficult to find serious ethical arguments for the kind of discrimination that is involved in giving less weight to future generations' well-being. Indeed, Hara *et al.* (2008) remind us of a highly relevant passage from Sidgwick (1907, p. 414):

> "How far we are to consider the interests of posterity when they seem to conflict with those of existing human beings? It seems, however, clear that the time at which a man exists cannot affect the value of his happiness from a universal point of view; and that the interests of posterity must concern a Utilitarian as much as those of his contemporaries, ..."

In his celebrated paper on optimal saving, Ramsey (1928, p. 261) famously started out by following the spirit of Sidgwick when he refused to discount the welfare of future generations:

> One point should perhaps be emphasised more particularly; it is assumed that we do not discount later enjoyments in comparison with earlier ones, a practice which is ethically indefensible and arises merely from the weakness of the imagination.[10]

Arrow (1999b) not only quotes this passage from Ramsey, but also adds two later quotes from other English economists. The first is from Pigou (1932, p. 25) stating that pure time preference "implies . . . our telescopic

[10] The latter part of the paper "no longer reckon[s] future utilities and disutilities as equal to present ones, but discount[s] them at a constant rate ρ." (Ramsey, 1928, p. 553).

faculty is defective." The second is Harrod's (1948, p. 40) claim that "[P]ure time preference [is] a polite expression for rapacity and the conquest of reason by passion."[11]

2.4 Discounting Our Own Future

When individuals contemplate their own future consumption, a discussion of how to discount it very like that in Section 2.2 might easily arise. On the other hand, if the same individuals contemplate their own future standard of living, there may be a closer parallel with Section 2.3. For example, an individual saving for retirement might want to discount future income somewhat if there is no longer any need to cover the expense of travelling to work, or of being able to live very close to work.

Such discounting of our own futures is quite different from what concerns us here. We are examining the ethical issue of whether there is any justification for discounting a life simply and only on the grounds that it starts later. It is not clear why the impatience of an individual who may value the future less than the present should be at all relevant to ethics. The fact some people are sometimes impatient in their own decisions does not tell us that there is any moral justification for discriminating between different people just because some are born decades later than others.

At this point it may be worth reminding the reader that market prices and interest rates, or rates of return, are very unlikely to give us ethical evaluations of the kind needed to guide society toward good decisions. Instead, they describe facts concerning the outcomes (equilibrium or otherwise) of the individual choices of many market participants. Indeed, market interest rates typically do not even give ethically appropriate individual marginal valuations, especially given the many interrelated imperfections that seem inevitable in capital markets — see, for example, Hammond (1992). Looking at market rates is rarely an ethically defensible route to the social evaluations that are necessary here.

We do mention, however, that just as mortality is one reason for discounting our own futures, so the possibility of human extinction is a reason for discounting social outcomes, as discussed in Section 8.

[11]Among other works dicussing intergenerational equity we mention Arrow *et al.* (1997) as well as Svensson (1980), Brown and Lewis (1981), Lauwers (1993, 1997a, b, c, 2010, 2017), Broome (1994), Atkinson (2001), Fleurbaey and Michel (2003), Basu and Mitra (2003, 2007a, b), Ponthière (2003), Bossert, Sprumont and Suzumura (2007), Mitra and Basu (2007), Zame (2007), Hara *et al.* (2008), Creedy and Guest (2008), Asheim (2010), Bossert and Suzumura (2011), Roemer (2011), Mertens and Rubinchik (2012), Quiggin (2012), Dubey and Mitra (2013).

2.5 The Strong Argument

The first reason that KJA gave for discounting, impatience, or time perspective was what we will call the "strong argument". This is to be distinguished from what Arrow (1999b) explicitly describes as the "weak Koopmans argument", which receives brief attention in Section 2.6 below.

To quote Arrow (1999b):[12]

> Why then not embrace the idea of zero time perspective? Koopmans in several classic papers (1960, 1964) gave a crushing answer; see also Brown and Lewis (1981) for a more general treatment. The argument seems recondite. Koopmans considers a world which lasts forever. Therefore choice (including ethically-based choice) is based on a preference ordering over infinite-dimensional consumption streams. He argues that if the ordering is continuous and also sensitive (i.e., if one stream is never worse than another and is better at one or more time points, then it must be strictly preferred), it must display impatience.
>
> A simple restatement of his reasoning can bring out the essential point. I confine myself to the intertemporally separable case. Imagine initially that output consists of a constant stream of completely perishable goods. There can be no investment by definition. Now imagine that an investment opportunity occurs, available only to the first generation. For each unit sacrificed by them, a perpetual stream of α per unit time is generated. If there were no time preference, ... we can say that given any investment, short of the entire income, a still greater investment would be preferred.

Thus, Arrow concludes that without impatience the optimal saving rate could become arbitrarily close to 100%. A similar conclusion emerges from the cake-eating example described by Gale (1967, p. 4, Example 2).

Nevertheless, the following passage from Koopmans (1960, pp. 287–288) suggests that he at least intended his results to be applied only in the rather different context of consumer choice:

> This study started out as an attempt to formulate postulates permitting a sharp definition of *impatience*, the short term Irving Fisher has introduced for preference for advanced timing of satisfaction. To avoid complications connected with the advancing

[12]The opening question in the quotation is a valuable addition in Arrow (1999b) to the corresponding passage in Arrow (1999a).

> age and finite life span of the individual consumer, these postulates were set up for a (continuous) utility function of a consumption program extending over an infinite future period. The surprising result was that only a slight strengthening of the continuity postulate . . . permits one to conclude from the existence of a utility function satisfying the postulates, that impatience prevails at least in certain areas of the program space.

Thus, it seems that Koopmans (and Diamond) started out by considering only consumers who discount their own future selves, as discussed in Section 2.4.[13] Nevertheless, it was natural for Arrow to consider the obvious extension to social choice theory, with an infinite series of successive generations. Indeed, this follows the tradition of the later works by Koopmans (1965, 1967), who even devotes part of these surveys to the Ramsey case when the welfare of future generations remains undiscounted. Then the same mathematical analysis which, under some conditions, shows that a consumer cannot treat equally consumption in an infinite number of periods, also rules out intergenerational equity, in the sense of treating all future generations equally.

2.6 The Weak Argument

Arrow (1999a) offered an additional reason for abandoning intergenerational equity:

> I therefore conclude that the strong ethical requirement that all generations be treated alike, itself reasonable, contradicts a very strong intuition that it is not morally acceptable to demand excessively high savings rates of any one generation, or even of every generation.

A very similar argument is adduced in Arrow (1999b), where he adds:

> Not merely is saving arbitrarily close to 100% unacceptable but very high sacrifices are also. I call this the *weak Koopmans argument.*

[13] Stern (2014b, p. 472; 2015, p. 169) also quotes a recent personal communication in which Peter Diamond, a co-author of Koopmans *et al.* (1964), had argued to the effect that the results of this line of work, if they would indeed preclude intergenerational equity, should not be applied to the issue of whether to discount the welfare of future generations. Instead Diamond has argued in favour of the kind of "pragmatic" criteria that we discuss near the end of Section 6.7.

And in Arrow (2007, p. 4) he writes:

> Tjalling Koopmans pointed out in effect that the savings rates implied by zero time preference are very much higher than those we observe. (I am myself convinced by this argument.)

The later joint paper by Arrow *et al.* (2013) even advocates a declining discount factor, though this may be justified if it is applied to monetary measures of consumer benefit when these are increasing over time.

Moreover, the concluding parts of Arrow (1999a, b) move toward considering the kind of growth framework with equilibrium in a dynamic game, as propounded by the foundational work of Phelps and Pollak (1968), which Arrow cites, as well as by Inagaki (1970, 1973), Dasgupta (1974, 1994), and others. Of course, the sophisticated — or what we would now call the "subgame perfect" equilibrium — outcome of such a game typically involves later plans that deviate from earlier intentions, as discussed in Hammond (1976b) and many subsequent writings.

This weak argument was discussed in Stern (2006, 2008, 2015) — see especially Stern (2008, p. 16). In his comment on the *Stern Report*, however, Arrow (2007, p. 4) himself suggests that, at least in the context of mitigating climate change, the discounting issue may lack practical importance:

> Many have complained about the Stern Review adopting a value of zero for ρ, the social rate of time preference. However, I find that the case for intervention to keep CO_2 levels within bounds (say, aiming to stabilize them at about 550 ppm) is sufficiently strong as to be insensitive to the arguments about ρ.[14]

Thus, KJA's reasoning included the recognition that, even with substantial pure-time discounting, unmanaged climate change has the potential to cause damage severe enough to justify strong action.

We would agree, while noting that Stern (2015) in particular discusses how many current economic models fail to capture adequately the immense potential damage. We also note this observation from Stern (2008, p. 19):

> . . . for any given set of structural risks and a utility function, pure time discounting . . . can be set so that the estimated damages are as small as we please.

[14]Since KJA wrote this, the scientific consensus on the potential damage due to climate change has become increasingly worrying. So the 550 ppm concentration level that KJA suggests should now be seen as too high. In the light of Stern (2008, p. 5), for example, that level seems incompatible with the Paris COP 21 aspiration that global warming since the pre-industrial era should not exceed 2° C. — or ideally, 1.5° C. Presumably this revision strengthens KJA's argument.

Actually the extraordinary technical progress in low and even negative carbon technologies since 2007, when KJA made that remark, suggest that his conclusion would be even more robust now.

3 Basic Framework: Consequences and Utilities

3.1 Consequentialist Rationality in Ethical Decision-Making

Following the early work of Keynes, Ramsey and de Finetti on subjective probability, Savage (1954) set out a path-breaking theory of rational choice. This was based on defining an act as a mapping from uncertain states of the world to consequences. Earlier Arrow (1951, p. 404), in his article on "choice in risk-taking situations", had summed up a key argument as follows:

> Among the actions actually available, then, that action is chosen whose consequences are preferred to those of any other available action.

Later, in the concluding part of Arrow (1963) — the definitive second edition of his Ph.D. dissertation that included several significant revisions — KJA addressed the question of why society should maximize a (complete and transitive) preference ordering. This inspired PJH to embark on a line of research that, among other things, derived the existence of a complete and transitive preference ordering from three "consequentialist hypotheses". These require that behaviour should not only be well-defined and dynamically consistent on an unrestricted domain of finite decision trees, but that its consequences should be explicable as the result of planned choices.[15]

Meanwhile PJH learned that "consequentialism" was a neologism that Anscombe (1958) had applied to an ethical doctrine she wanted to criticize. Indeed, to quote the entry on Anscombe in the online *Stanford Encyclopedia of Philosophy*, this article of hers "stimulated the development of virtue ethics as an alternative to Utilitarianism, Kantian Ethics, and Social Contract theories." Yet, as is characteristic of his broad scholarship, KJA had earlier introduced PJH to St. Thomas Aquinas's much more authoritative rejection of consequentialism:

[15] Relevant writings include Hammond (1986) in the Arrow *Festschrift*, as well as Hammond (1996a), which was an invited contribution to a conference that KJA co-organized during his presidency of the International Economic Association.

> On the Contrary, The consequences do not make an action evil that was good; nor good one that was evil.[16]

Anscombe's critique of consequentialism appears to arise from a notion of "consequence" that was too limited to include the kind of virtues that she regarded as important. The modern school of virtue ethics that she inspired harks back to concepts from the ancient Greeks such as *eudaimonia*, often translated as "human flourishing". Though even here, Anscombe (1958, p. 15) issues the following warning

> . . . philosophically there is a huge gap, at present unfillable as far as we are concerned, which needs to be filled by an account of human nature, human action, the type of characteristic a virtue is, and above all of human "flourishing". And it is the last concept that appears the most doubtful.

3.2 Consequentialization

Despite Anscombe's work, it is clear that in principle consequences should be defined sufficiently broadly to include everything, even *eudaimonia*, that philosophers can legitimately claim to be ethically relevant. Other relevant parts of any consequence could include considerations that would lead Aquinas to judge the goodness or evil of an act, as well as issues of concern to the "impartial spectator" who plays such a key role in Adam Smith's *Theory of Moral Sentiments*.[17]

Indeed, the possibility of this process of "consequentialization" has been explored by several philosophers, including Portmore (2007, 2009), Brown (2011), and Mukerji (2016). One must admit, however, that as the concept of bounded rationality recognizes, fruitful analysis of any practical ethical decision problem will most likely require a considerably less unwieldy concept of consequence. Yet in the end any coherent consequentialist theory of

[16] From *Summa Theologiae*, First Part of the Second Part, Question 20, Article 5, Objection. The Latin original is "Sed contra, eventus sequens non facit actum malum qui erat bonus, nec bonum qui erat malus." KJA recalled a version due to Dorothy Sayers in the commentary to her English translation of Dante Alighieri's *Divina Commedia*.

[17] This paragraph is inspired in part by the issues that Patrick Suppes kindly raised when discussing Hammond (1988) at the May 1986 conference on "Distributive Justice and Inequality" at the Wissenschaftskolleg zu Berlin. At the time, the most meaningful part of PJH's inadequate response may have been the remark that philosophers like Pat excel at drawing attention to the ethical relevance of consequences that belong to domains much richer than those usually considered by economists.

ethical decision making, even one that allows for bounded rationality, presumably must be based on a suitably specified consequence domain, even if that domain must be modelled incompletely.

3.3 A Universal Domain of Personal Consequences

Our interest is in applying prescriptive social choice theory to an issue that arises in welfare economics — namely, specifying what discount rates one should apply to future generations' welfare. For this one uses an individualistic theory of social consequences that starts with a "universal" *personal consequence domain*, which we take to be a non-empty set Y whose typical member y has many attributes or dimensions. Following the discussion of Sections 3.1 and 3.2, we postulate that each $y \in Y$ is "all-inclusive" in the sense that includes everything that is ethically relevant to any decision, whether individual or social, that concerns a person's life history. These dimensions should include the individual's own preferences and beliefs, insofar as they are deemed relevant. They should also allow for variations in the date and circumstances surrounding an individual's birth, upbringing, and death, including those aspects that also affect parents, partners and family members. Thus, with one exception, we assume that each consequence $y \in Y$ has attributes which include personal copies of any ethically relevant common or impersonal circumstances that are shared with other persons.

The one exception, which we use repeatedly from Section 5 on when discussing demographic consequences, is that we postulate one particular *non-existence* consequence $y_0 \in Y$. This is the unique personal consequence that comes about if and only if the person concerned never exists.

3.4 Social Consequences as Personal Consequence Profiles

Initially, we consider a finite fixed potential population of size $n \in \mathbb{N}$. Thus, the set of numerical labels is taken to be

$$\mathbb{N}_n := \{1, 2, \ldots, n\} = \{i \in \mathbb{N} \mid i \leq n\} \tag{1}$$

Later, starting in Section 6, we will extend our analysis to potentially infinite populations. This is of course the case that was of most concern to Ramsey and Arrow. For now, however, with population fixed at size n, we take the *social consequence domain* to be the n-fold Cartesian product Y^n of the personal consequence domain Y. So each social consequence $y^n \in Y^n$ is an n-dimensional profile

$$y^n = \langle y_i \rangle_{i=1}^n = \langle y_i \rangle_{i \in \mathbb{N}_n} \tag{2}$$

of personal consequences for the n individuals who are given numbers in the set $\mathbb{N}_n \subset \mathbb{N}$ defined by (1).

3.5 Biased and Extended Original Positions

Vickrey (1945, 1960) and Harsanyi (1953, 1955, 1977, 1978, 1979) independently formulated the idea that ethical social decisions would be those that were taken impartially in a version of what Rawls (1971) later described as an "original position", behind a "veil of ignorance" where the decision-maker does not know which person she or he will become eventually. Of course, Rawls famously suggested that such decisions should satisfy his "difference principle", which requires that any differences in individuals' allocations of "primary goods" had to be justified because they would make all individuals better off. This gave rise to the "Rawlsian maximin" rule that would maximize the minimum level of well-being. It was a principle that Harsanyi (1975) in particular criticized. By contrast, what we will call the *Vickrey–Harsanyi original position* requires the impartial ethical decision-making agency to contemplate what it would choose in case it faced an even chance lottery whose different outcomes were the personal consequences of the various individuals in society — see also the discussion by Mongin (2001) and others of Harsanyi's "impartial observer".

The theory presented here will also accommodate lotteries in the form of *biased original positions* where, upon emerging from behind the probabilistic version of the veil of ignorance, there are arbitrary specified probabilities μ_i of becoming different people $i \in \mathbb{N}_n$. Indeed, we consider *extended original positions* where these biased probabilities of becoming different people can even be chosen.

To include these extended original position lotteries, we start with a general domain of ordinary finite decision trees in which the consequence attached to each terminal node is a lottery $\lambda \in \Delta(Y^n)$ over the domain Y^n of social consequences. To include any biased original position, we replace each such terminal node where the consequence lottery is $\lambda \in \Delta(Y^n)$ with some "original position" chance node. At any such node, there is a specified probability distribution $\mu \in \Delta(\mathbb{N}_n)$ that gives the probability μ_i of becoming each individual numbered $i \in \mathbb{N}_n$. Thereafter, person i experiences the relevant marginal consequence lottery $\operatorname{marg}_i \lambda$ over i's personal copy of the domain Y of extended personal consequences. Given $\lambda \in \Delta(Y^n)$, this marginal lottery has probabilities given by

$$\lambda_i(y) := \operatorname{marg}_i \lambda(y) := \lambda(\{y^n \in Y^n \mid y_i = y\}) \tag{3}$$

3.6 Fundamental Cardinal Utility

We follow the standard economist's view that ethical theory should prescribe normatively appropriate decisions for society. Specifically, following Hammond (1996a, 1998), we impose rationality and continuity postulates requiring that prescribed social behaviour should:

1. be well-defined over the domain of decision nodes in all finite decision trees with consequences in the domain $\Delta(Y^n)$ of lotteries over social consequences, excluding only decision trees in which one or more zero probabilities are attached at chance nodes of the tree;

2. determine consequences that are explicable as the planned choice of ethically desirable lotteries over social consequences;

3. satisfy "dynamic consistency" in the sense that behaviour prescribed at a decision node in any continuation subtree should match what had been prescribed earlier for that node;

4. vary continuously w.r.t. the non-zero probabilities at any chance node of a tree.

As explained in many works, these four postulates together imply the existence of a cardinal equivalence class of von Neumann–Morgenstern utility functions $Y \ni y \mapsto u(y) \in \mathbb{R}$ with the property that the consequences of prescribed behaviour in any finite decision tree should be a consequence lottery $\lambda \in \Delta(Y)$ that maximizes lifetime expected utility

$$\mathbb{E}_\lambda u = \sum\nolimits_{y \in Y} \lambda(y)\, u(y) \tag{4}$$

over the finite set of lifetime consequence lotteries that are feasible in the tree. Here we are using the familiar definition that any two utility functions $y \mapsto u(y)$ and $y \mapsto \tilde{u}(y)$ are *cardinally equivalent* just in case there exist an additive constant $\alpha \in \mathbb{R}$ and a *positive* mulitplicative constant $\rho \in \mathbb{R}$ such that

$$\tilde{u}(y) = \alpha + \rho u(y) \text{ for all } y \in Y \tag{5}$$

Thus, the two utility functions $y \mapsto u(y)$ and $y \mapsto \tilde{u}(y)$ are cardinally equivalent if and only if, for every triple $\{a, b, c\}$ of consequences in Y, the ratios of utility differences satisfy

$$\frac{\tilde{u}(a) - \tilde{u}(c)}{\tilde{u}(b) - \tilde{u}(c)} = \frac{u(a) - u(c)}{u(b) - u(c)} \tag{6}$$

whenever both denominators are non-zero because b and c are not indifferent.

3.7 Interpreting Ratios of Utility Differences

Consider the Marschak triangle whose three vertices are the three degenerate lotteries δ_y ($y \in \{a, b, c\}$) in which the particular consequence y occurs with probability 1. Thus, each point of the triangle corresponds to a unique lottery $\lambda \in \Delta(Y)$ with the property that

$$\lambda(\{a, b, c\}) = \lambda(a) + \lambda(b) + \lambda(c) = 1$$

As discussed in Hammond (1998, Section 2.3), the ratios on each side of equation (6) represent the constant slopes of the parallel straight line indifference curves in this triangle, along which expected utility

$$\mathbb{E}_\lambda u = \lambda(a)\, u(a) + \lambda(b)\, u(b) + \lambda(c)\, u(c)$$

is constant. Indeed, the slopes on each side of equation (6) equal the constant marginal rate of substitution (MRS) between:

1. any increase in the probability $\lambda(a)$ of getting a that is offset by an equal decrease in the probability $\lambda(c)$ of getting c;
2. any increase in the probability $\lambda(b)$ of getting b that is offset by an equal decrease in the probability $\lambda(c)$ of getting c.

Notice that all variations in individuals' consequences, including general characteristics that affect their welfare and even their existence, are included in the extended personal consequence y. Thus, despite the objections of Broome (1993), we are treating any "cause of preference" — in the sense of any personal characteristic that affects our ethical judgements of a person's wellbeing — as an "object of preference", including those preferences that can be regarded as interpersonal comparisons. In fact, any utility function $Y \ni y \mapsto u(y) \in \mathbb{R}$ in the cardinal equivalence class of von Neumann–Morgenstern utility functions represents what Tinbergen (1957) and Kolm (1994) call "fundamental preferences" over the domain Y.[18] Indeed, we go further, and require the expected utility function

$$\Delta(Y) \ni \lambda \mapsto \mathbb{E}_\lambda u = \sum\nolimits_{y \in Y} \lambda(y)\, u(y) \tag{7}$$

to represent fundamental preferences over the domain $\Delta(Y)$ of lotteries λ.

[18] Arrow (1977) himself explores a similar concept.

3.8 General Discrete Lotteries and Bounded Utility

Menger (1934) showed how to modify the well-known St. Petersburg paradox so that it applies to any unbounded utility function. This result led Arrow (1951; 1965, pp. 28–44; 1971, ch. 2; 1972) to insist that utility should be bounded. Indeed, suppose that expected utility is to be extended to a continuous function defined not only over simple lotteries whose support is a finite set of possible outcomes, but also over general discrete lotteries whose support is a countably infinite set of possible outcomes, including extinction dates. Then analysis such as that in Hammond (1998, Section 8) shows that relevant marginal rates of substitution like (6) must be bounded away from both zero and infinity.

Accordingly, from now on we make the following assumption:

Assumption 1. *There exist both a common lower bound* $\underline{u}$ *and a common upper bound* $\bar{u} > \underline{u}$ *such that, for all possible consequences* $y \in Y$ *that any potential individuals may face, one has*

$$\underline{u} \leq u(y) \leq \bar{u} \tag{8}$$

4 Basic Framework: Suppes Equity

4.1 Expected Utility in a Biased Original Position

Consider any biased original position in which μ_i denotes the probability of becoming the person bearing the numerical label $i \in \mathbb{N}_n$. The fundamental utility function u on Y has been constructed so that, in case there are n individuals, the ethical decision maker should choose a consequence lottery $\lambda \in \Delta(Y^n)$ in order to maximize overall expected utility.

In the biased original position described by $\mu \in \Delta(\mathbb{N}_n)$, the relevant expected utility calculation involves a simple lottery on the Cartesian product set $Y \times \mathbb{N}_n$ whose members (y, i) pair a personal consequence y with a numerical label i. The relevant lottery, which we denote by $\mu \circ \lambda$, comes from compounding λ with μ. For every (y, i) in the domain $Y \times \mathbb{N}_n$, the relevant compound probability is given by

$$(\mu \circ \lambda)(y, i) := \mu_i \cdot \operatorname{marg}_i \lambda(y) = \mu_i \cdot \lambda_i(y) \tag{9}$$

Here, for each $i \in \mathbb{N}_n$, we have followed (3) in using λ_i to denote the marginal probability distribution $\operatorname{marg}_i \lambda$ of i's random personal consequence.

Now we can calculate the expected utility derived from a chosen biased original position $\mu \in \Delta(\mathbb{N}_n)$ combined with a chosen lottery consequence $\lambda \in \Delta(Y^n)$ as

$$\mathbb{E}_{\mu\circ\lambda}[u(y)] := \sum\nolimits_{(y,i)\in Y\times\mathbb{N}_n} (\mu \circ \lambda)(y, i)\, u(y) \tag{10}$$

Finally, we introduce a simplification that comes from realizing that the joint probability distribution $\mu \circ \lambda$ on $Y \times \mathbb{N}_n$ is equivalent to a two-stage lottery where first $i \in \mathbb{N}_n$ is determined before i's personal consequence $y_i \in Y$. This leads to an alternative iterated expected value

$$\mathbb{E}_{\mu\circ\lambda}[u(y)] = \mathbb{E}_{\mu}\left[\mathbb{E}_{\lambda_i}[u(y_i)]\right] = \sum\nolimits_{i\in\mathbb{N}_n} \mu_i \sum\nolimits_{y_i\in Y} \lambda_i(y_i)\, u(y_i) \tag{11}$$

of individual expected utilities arising from the relevant personal consequence lotteries λ_i.

4.2 An Original Position with Suppes Equity

In practice, not even our idealized ethical decision-making agency can choose what individual it will become upon emerging from behind the veil of ignorance. Thus, it must treat the biased original position $\mu \in \Delta(\mathbb{N}_n)$ as fixed. Then it is reduced to choosing $\lambda \in \Delta(Y^n)$ in order to maximize the function (11) while treating $\mu \in \Delta(\mathbb{N}_n)$ as a fixed vector of probabilistic weights. So the appropriate objective is the weighted utilitarian *Bergson social welfare function* (or BSWF) defined by

$$\Delta(Y^n) \ni \lambda \mapsto W(\lambda; \mu) := \mathbb{E}_{\mu\circ\lambda}[u(y)] \tag{12}$$

For "two-person decision situations" Suppes (1966, definition 5, p. 296) defines a preference relation that he calls "more just than". Sen (1970, Chapter 9*) first offers an extension of Suppes' definition from 2 persons to n. Then, after introducing interpersonal comparisons of utility levels into Suppes' framework, he relates this extended principle to Rawls' difference principle, and the corresponding maximin preference ordering. This relation was discussed further in Hammond (1976, 1979) and in many subsequent works by other social choice theorists.

Applied to the weighted utilitarian Bergson social welfare function (12), a pair of lotteries $\nu, \rho \in \Delta(Y^n)$ constitutes a *two-person decision situation* just in case there are two individuals $j, k \in \mathbb{N}_n$ such that for all other persons $i \in \mathbb{N}_n \setminus \{j, k\}$, the marginal distributions $\nu_i, \rho_i \in \Delta(Y)$ satisfy $\nu_i = \rho_i$. In this case equations (12) and (11) imply that the welfare difference between ν and ρ is

$$W(\nu;\mu) - W(\rho;\mu) = \mu_j \sum\nolimits_{y_j \in Y} [\nu_j(y_j) - \rho_j(y_j)]\, u(y_j) \\ + \mu_k \sum\nolimits_{y_k \in Y} [\nu_k(y_k) - \rho_k(y_k)]\, u(y_k) \quad (13)$$

Now, *Suppes equity* insists that interchanging the pairs of marginal lotteries ν_j, ν_k and ρ_j, ρ_k of these two individuals should have no effect on the social preference between ν and ρ in $\Delta(Y^n)$. In other words, we should consider the new lotteries $\tilde{\nu}, \tilde{\rho} \in \Delta(Y^n)$ whose marginal distributions $\tilde{\nu}_i, \tilde{\rho}_i \in \Delta(Y)$ satisfy $\tilde{\nu}_i = \nu_{\tau^{j,k}(i)}$ and $\tilde{\rho}_i = \rho_{\tau^{j,k}(i)}$ for all $i \in \mathbb{N}_n$, where $\mathbb{N}_n \ni i \mapsto \tau^{j,k}(i) \in \mathbb{N}_n$ is the transposition mapping that interchanges individuals j and k while leaving all other individuals unaffected. Then Suppes equity requires that

$$W(\tilde{\nu};\mu) \gtreqless W(\tilde{\rho};\mu) \text{ according as } W(\nu;\mu) \gtreqless W(\rho;\mu) \quad (14)$$

Evidently (13) is consistent with (14) for all pairs $\nu, \rho \in \Delta(Y^n)$ if and only if $\mu_j = \mu_k$. Given that $\mu \in \Delta(\mathbb{N}_n)$ is a probability distribution, evidently Suppes equity holds for all pairs of individuals $j, k \in \mathbb{N}_n$ if and only if the weights satisfy $\mu_i = \frac{1}{n}$ for all $i \in \mathbb{N}_n$. So instead of the weighted sum (11), Suppes equity implies that we should have an *unweighted* utilitarian BSWF of the form

$$\Delta(Y^n) \ni \lambda \mapsto W(\lambda) := \frac{1}{n} \sum\nolimits_{i \in \mathbb{N}_n} \sum\nolimits_{y_i \in Y} \lambda_i(y_i)\, u(y_i) \quad (15)$$

This, of course, is precisely the form that Vickrey (1945, 1960) and Harsanyi (1953, 1955, 1977, 1978, 1979) advocated, taking the view that the original position requires the ethical decision maker to be impartial or unbiased in the sense that all individuals should be given equal weight.

5 Basic Framework: Demographic Consequences

5.1 Variable Numbers and Possible Non-Existence

Up to now we have considered what Parfit (1984) called "same numbers" problems, where the population of individuals is fixed. To allow for the probability of extinction, however, we must discuss "variable numbers" problems where the population varies, and so there are variable "demographic" consequences.

To accommodate these, we follow Hammond (1988) and also Blackorby, Bossert and Donaldson (2005) (henceforth BBD) in assuming that all decisions which only affect the utility of a non-existent individual should be regarded as indifferent. In fact, in Section 3.3 we already introduced the assumption that there is one particular personal consequence $y_0 \in Y$ which comes about if and only if the person concerned never comes into existence. Let $u_0 := u(y_0) \in \mathbb{R}$ denote the utility level associated with non-existence.

5.2 Avoiding Parfit's "Repugnant Conclusion", I

At this point BBD choose a positive *critical level* of utility that we denote by u_C. This is defined so that, rather than an individual not living at all, it is better or worse for the society that an individual should live and experience personal consequence $y \in Y$ according as $u(y) \gtrless u_C$. For BBD, the alternative criterion $u(y) \gtrless 0$ indicates whether a person who actually comes into existence and experiences personal consequence y would (or should) personally prefer or disprefer that life to not living at all.

One motivation that BBD give for introducing the positive constant u_C is to avoid the "repugnant conclusion" enunciated by Parfit (1984). This involves the observation that, in a society of n individuals who all experience an identical personal consequence y, a total utility social welfare function would be

$$\mathbb{N} \times Y \ni (n, y) \mapsto S(n, y) := n\, u(y) \tag{16}$$

Such an objective implies that, whenever $u(y) > 0$, it is always better to increase n with $u(y)$ fixed. Following Parfit, BBD see this as "repugnant" when $u(y)$ is close to zero, which is interpreted as indicating a miserable personal consequence. Their positive critical level u_C allows BBD to replace the objective function (16) with the alternative

$$\mathbb{N} \times Y \ni (n, y) \mapsto S_0(n, y) := n\,[u(y) - u_C] \tag{17}$$

By choosing u_C far enough above the zero level at which an individual is thought to be on the margin of preferring not to have be born, the functional form (17) enables BBD to escape the repugnant conclusion.

5.3 Avoiding Parfit's "Repugnant Conclusion", II

By contrast, following Hammond (1988, 1991) and Fleurbaey and Hammond (2004), we normalize the common interpersonally comparable utility function $Y \ni y \mapsto u(y) \in \mathbb{R}$ so that

$$u(y_0) = 0 \tag{18}$$

That is, the critical level is set to zero.

Now consider any fixed personal consequence profile y^n as in (2) which occurs with probability 1. The welfare function defined by (15) takes the form

$$\frac{1}{n} \sum\nolimits_{i \in \mathbb{N}_n} u(y_i) \tag{19}$$

But the set of individuals who ever exist in this profile is

$$L(y^n) := \{ i \in \mathbb{N}_n \mid y_i \in Y \setminus \{y_0\} \} \tag{20}$$

So the normalization (18) obviously reduces (19) to the average utility

$$\frac{1}{n} \sum\nolimits_{i \in L(y^n)} u(y_i) \tag{21}$$

over all n individuals, whether or not they exist. This has important implications later — see especially Section 5.7.

To avoid the repugnant conclusion implied by maximizing (16) when the constant u_C in (17) is zero, we simply need to extend the argument of Hammond (1988) concerning how to define the zero level of utility. Specifically, we admit the possibility of personal consequences $y \in Y$ that, if they occurred with certainty, would be so repugnant that any ethical agent like Adam Smith's impartial spectator would prefer, *ceteris paribus*, that the individual never exist. This is entirely consistent with the utility function $y \mapsto u(y)$ having the property that, for each lottery $\lambda \in \Delta(Y)$, one has $\sum_{y \in Y} \lambda(y)\, u(y) \gtreqless 0$ according as the society as a whole is made better or worse off, in the view of the ethical agent, by adding an extra individual whose random personal consequence *ex ante* is given by λ. For any personal consequence satisfying $u(y) < 0$, the absolute value $|u(y)|$ measures the repugnance that the impartial spectator should feel at seeing any individual condemned to such a miserable existence.

5.4 Normalized Individual Welfare

Following the terminology of Roberts (1980), it follows that utility is now measured up to a *cardinal ratio scale*. This means that the two utility functions $y \mapsto u(y)$ and $y \mapsto \tilde{u}(y)$ are equivalent just in case there exists a *positive* multiplicative constant $\rho \in \mathbb{R}$ such that $\tilde{u}(y) = \rho\, u(y)$ for all $y \in Y$. This first normalization excludes additive shifts represented by a non-zero α in (5).

From now on we ignore the trivial case when $u(y) = 0$ for all $y \in Y$, which would imply that all personal consequences are regarded as entirely indifferent to the personal consequence y_0 associated with not even existing. Indeed, it is surely reasonable to assume that there exists at least one favourable personal consequence $y^* \neq y_0$ such that $u(y^*) > 0$ because ethically it is deemed better to experience y^* than not to live at all.

5.5 Interpreting Interpersonal Comparisons

Arrow (1977) offered an interpretation of interpersonal comparisons of utility levels of individual welfare. Indeed, the title he chose there links the paper to the earlier work in Arrow (1951) on the concept of "extended sympathy", though that link is left largely implicit. Here, we have utility measured on a cardinal ratio scale that gives decision-theoretic meaning not only to level comparisions of utility, but also to ratios of non-zero utilities. Indeed, as discussed in Hammond (1991) and Fleurbaey and Hammond (2004), a comparison $u(y) > u(y')$ means that experiencing personal consequence y is better than experiencing personal consequence y', even if these personal consequences are experienced by different individuals. Again, this ethical opinion may be that of an impartial spectator.

On the other hand, because of the normalization $u(y_0) = 0$, given any two personal consequences $y, y' \in Y \setminus \{y_0\}$, the well-defined numerical ratio $u(y)/u(y')$ must satisfy

$$\frac{u(y)}{u(y')} = \frac{u(y) - u(y_0)}{u(y') - u(y_0)} \tag{22}$$

Then, following the discussion following (6) in Section 3.6, this ratio of utility differences must equal the constant marginal rate of substitution between any two different net probability shifts that both move probability by the same absolute amount:

1. away from the non-existence consequence y_0;

2. toward personal consequence y, as opposed to y'.

It may bear repeating yet again that this constant marginal rate of substitution should be that of an impartial spectator.

5.6 Average versus Total Utilitarianism

Classical utilitarianism advocates maximizing the expected value of *total* utility, given by the *classical* BSWF defined by

$$\Delta(Y^n) \ni \lambda \mapsto W^{\text{total}}(\lambda) := \sum\nolimits_{i \in \mathbb{N}_n} \sum\nolimits_{y_i \in Y} \lambda_i(y_i)\, u(y_i) \tag{23}$$

Of course, this objective was already advocated by Hutcheson (1725), who famously wrote: "That Action is best, which procures the greatest Happiness for the greatest Numbers".

By contrast, the alternative BSWF (15), where $W^{\text{total}}(\lambda)$ gets divided by the number of individuals n, looks like *average utilitarianism*, apparently first advocated by Edgeworth (1925).[19] We note, however, that Sidgwick (1907, pp. 415–416) wrote earlier as follows:

> ... strictly conceived, the point up to which, on Utilitarian principles, population ought to be encouraged to increase, is not that at which average happiness is the greatest possible ... but that at which the product formed by multiplying the number of persons living by the amount of average happiness reaches its maximum".

Evidently, when n is truly fixed, then (15) and (23) are cardinally equivalent functions of λ, so maximizing either is equivalent to maximizing the other. Yet when population is being affected by policy choices, it is usual to interpret this as choosing n.

As argued in Hammond (1988), in decision trees where population is being chosen, maximizing the average utilitarian criterion (15) will be dynamically consistent only if the chosen population size n is calculated after keeping track of all the individuals who ever existed, starting from a date no later than the beginning of the planning process. This requires that the number n never be updated. Far in the future, we will have to count the number of long-dead ancestors who should otherwise be entirely forgotten.

5.7 Average Utility in an Expanded Population

There is also a more direct argument that justifies the classical or total utilitarian criterion (23). Recall that we have defined n as the total number of *potential* individuals in each consequence profile $y^n \in Y^n$. These potential individuals include all those $i \in N_n$ who never come into existence in the personal consequence profile $y^n \in Y^n$ because their own personal consequence is y_0.

Consider now a population of size m which is much larger than n, and an expanded consequence domain $Y^n \times \{y_0\}^{m-n}$ of personal consequence

[19] See also Yaari (1981). In this connection, Harsanyi's work is generally interpreted as supporting average utilitarianism. The difference between classical and average utilitarianism also receives attention in Arrow, Dasgupta and Mäler (2003).

profiles $y^m = \langle y_i \rangle_{i=1}^m$ that, for all $i \in \mathbb{N}_m \setminus \mathbb{N}_n = \{n+1, n+2, \dots, m-1, m\}$, satisfy $y_i = y_0$ and so $u(y_i) = 0$.

In the Suppes version of the original position, where there is a probability $1/m$ of becoming any individual, the expected welfare of the social consequence $y^m \in Y^n \times \{y_0^{m-n}\}$ is

$$W_m^{\text{ave}}(y^m) := \frac{1}{m} \sum\nolimits_{i \in \mathbb{N}_m} u(y_i) = \frac{n}{m} \frac{1}{n} \sum\nolimits_{i \in \mathbb{N}_n} u(y_i) \tag{24}$$

where the last equality holds because $y_i = y_0$ and so $u(y_i) = 0$ for all $i \in \mathbb{N}_m \setminus \mathbb{N}_n$. Hence

$$W_m^{\text{ave}}(y^m) = \frac{n}{m} \frac{1}{n} \sum\nolimits_{i \in \mathbb{N}_n} u(y_i) = \frac{n}{m} W_n^{\text{ave}}(y^n) \tag{25}$$

where the last term is the product of:

- the probability n/m of being among the n individuals out of m who might ever exist;
- the conditional expected utility $W_n^{\text{ave}}(y^n)$ of being in an original position in a potential population consisting of those n individuals.

Multiplying each side of (25) by the positive constant m gives

$$m\, W_m^{\text{ave}}(y^m) = \sum\nolimits_{i \in \mathbb{N}_m} u(y_i) = \sum\nolimits_{i \in \mathbb{N}_n} u(y_i) = n\, W_n^{\text{ave}}(y^n) \tag{26}$$

and so

$$W_m^{\text{total}}(y^m) = \sum\nolimits_{i \in \mathbb{N}_m} u(y_i) = \sum\nolimits_{i \in \mathbb{N}_n} u(y_i) = W_n^{\text{total}}(y^n) \tag{27}$$

In particular, the first expression in (27) is independent of m provided that m is large enough. So, to be sure of including all potential individuals, we take the limit as $m \to \infty$. Then, on the understanding that $y_i = y_0$ and so $u(y_i) = 0$ for all $i > n$, an appropriate maximand is the *total* utilitarian welfare function

$$W_\infty^{\text{total}}(y^{\mathbb{N}}) := \sum\nolimits_{i=1}^{\infty} u(y_i) = \sum\nolimits_{i=1}^{n} u(y_i) \tag{28}$$

of the infinite personal consequence profile $y^{\mathbb{N}}$ that belongs to the Cartesian product $Y^{\mathbb{N}}$ of a countably infinite set of copies of Y.

Note that in the Suppes original position, as n increases, so does the probability n/m of coming into existence, no matter how large m may be. This is reflected in the total utilitarian objective (28), but not in the average utilitarian objective

$$W_n^{\text{ave}}(y^{\mathbb{N}}) = \frac{1}{n} \sum\nolimits_{i=1}^{n} u(y_i) \tag{29}$$

where n is being chosen.

6 Toward Intergenerational Equity

6.1 Unbounded Population

Up to now, we have really considered only finite-dimensional personal consequence profiles of the form $y^m \in Y^m$, or the derived utility streams

$$u^m = u^m(y^m) = \langle u(y_i) \rangle_{i=1}^m \in \mathbb{R}^m \tag{30}$$

Here m is a finite upper bound on the number of individuals who might ever exist.

At the end of Section 5.7, however, we did consider the total utilitarian sum (28) in which the infinite sum $\sum_{i=1}^{\infty} u(y_i)$ collapsed to the finite sum $\sum_{i=1}^{n} u(y_i)$ because we took the case when, for all $i > n$, one has $y_i = y_0$ and so $u(y_i) = 0$. Yet the debate between Ramsey, Koopmans and Arrow (*inter alia*) revolves entirely around what to do when potentially there are infinitely many individuals, in infinitely many generations.

So from now on, instead of finite-dimensional personal consequence profiles, we consider infinite-dimensional personal consequence *sequences* or *streams*

$$y^{\mathbb{N}} = \langle y_i \rangle_{i \in \mathbb{N}} \in Y^{\mathbb{N}} = \prod\nolimits_{i \in \mathbb{N}} Y_i \tag{31}$$

Note that we are not presuming that an infinite set of individuals will come into existence. Rather, given any finite $m \in \mathbb{N}$, there could be a positive probability of a stream $y^{\mathbb{N}}$ for which there exists some $i > m$ for whom i could exist, and so $y_i \neq y_0$.

6.2 Generational Structures

The issue that concerned Arrow (1999a, b) was whether one should discount the lives of individuals who belong to future generations. To discuss this formally, we need to define the concept of generation in our framework. This involves recognizing that when individuals are born depends upon the personal consequence stream $y^{\mathbb{N}} \in Y^{\mathbb{N}}$, which is of course influenced by policy choices.

Accordingly, let T denote the set of possible times or dates. Like the labels of the set of potential individuals, it is taken to be a copy of $\mathbb{N}$, the countably infinite set of natural numbers. Corresponding to each time $t \in T$ and each consequence profile $y^{\mathbb{N}} \in Y^{\mathbb{N}}$, there is a *generation* $G_t(y^{\mathbb{N}}) \subset \mathbb{N}$, which may be thought of as the finite set of numerical labels of all individuals born at date t when the consequence profile is $y^{\mathbb{N}}$. Since birth implies coming into existence, one has

$$i \in \cup_{t \in T} G_t(y^{\mathbb{N}}) \Longrightarrow y_i \neq y_0 \tag{32}$$

Definition 1. *A* generational structure *is a sequence* $G^T = \langle G_t \rangle_{t \in T}$ *of pairwise disjoint finite subsets of* $\mathbb{N}$ *that satisfy:*

1. *either* $G_t \neq \emptyset$ *for all* $t \in T$*, or there exists a finite* horizon $H \in T$ *such that* $G_t \neq \emptyset \Longleftrightarrow t \leq H$*;*

2. *for each* $t \in T$ *such that* $G_t \neq \emptyset$*, there exists a pair* $a_t, b_t \in \mathbb{N}$ *such that* G_t *is the set* $\mathbb{N} \cap [a_t, b_t]$ *of* $b_t - a_t + 1$ *consecutive natural numbers, where* $b_0 = 0$ *and then, for* $t = 1, 2, 3, \ldots$*: (i) the first member of generation* G_t *is* $a_t = b_{t-1} + 1$*; (ii) the last member of generation* G_t *is* $b_t = a_{t+1} - 1$*.*

An implicit assumption here is that, after the first time t at which generation G_t has no members, all subsequent generations are also empty. This seems relatively harmless. The following result will be used later.

Lemma 1. *For every generational structure* G^T*:*

1. *for every* $s \in T$*, one has* $\cup_{t=1}^{s} G_t = \{1, 2, \ldots, b_s\} \subset \mathbb{N}$*;*

2. *if* $G_t \neq \emptyset$ *for all* $t \in T$*, then* $\cup_{t \in T} G_t = \mathbb{N}$*.*

Proof. Part 1 of Lemma 1 is obvious from Part 2 of Definition 1.

Also, because $G_t \subset \mathbb{N}$ for all $t \in T$, one must have $\cup_{t \in T} G_t \subseteq \mathbb{N}$.

Conversely, in case $G_t \neq \emptyset$ for all $t \in T$, for each $s \in T = \mathbb{N}$ one must have $s \leq \# \cup_{t=1}^{s} G_t = b_t$, by Part 1 of Lemma 1. But then Part 2 of Definition 1 evidently implies that $s \in \cup_{t=1}^{s} G_t \subset \cup_{t \in T} G_t$. This proves that $\mathbb{N} \subseteq \cup_{t \in T} G_t$. □

From now on, we assume that individual $i \in \mathbb{N}$ can exist in the personal consequence stream $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ if and only if i belongs to a generation $G_t(y^{\mathbb{N}})$. Furthermore, the date $t \in T$ at which individual $i \in \mathbb{N}$ is born, and so the generation $G_t(y^{\mathbb{N}})$ to which i belongs, can only be affected by the personal consequences of people who belong to a preceding generation. That is, there can be no backward causation in creating new generations. Formally:

Assumption 2. *For each consequence stream* $y^{\mathbb{N}} = \langle y_i \rangle_{i \in \mathbb{N}} \in Y^{\mathbb{N}}$ *there exists a unique generational structure* $G^T(y^{\mathbb{N}}) := \langle G_t(y^{\mathbb{N}}) \rangle_{t \in T}$ *with the property that* $i \in \cup_{t \in T} G_t(y^{\mathbb{N}}) \Longleftrightarrow y_i \neq y_0$ *and also* $G^T(y^{\mathbb{N}}) = \mathbb{N} \cap [a_t(y^{\mathbb{N}}), b_t(y^{\mathbb{N}})]$*.*

Moreover, given any pair $y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}} \in Y^{\mathbb{N}}$ *of alternative consequence profiles, if* $y_i = \tilde{y}_i$ *for all* $i \leq b_{t-1}(y^{\mathbb{N}}) = a_t(y^{\mathbb{N}}) - 1$*, then* $G_t(y^{\mathbb{N}}) = G_t(\tilde{y}^{\mathbb{N}})$*.*

6.3 Generational Welfare

To define the welfare of each generation G_t, we consider the domains of:

1. *one-generation situations* in the form of a pair $y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}} \in Y^{\mathbb{N}}$ of consequence streams for which there exists $t \in T$ such that:

   (a) $G_t(y^{\mathbb{N}}) = G_t(\tilde{y}^{\mathbb{N}})$;

   (b) for all $i \notin G_t(y^{\mathbb{N}}) = G_t(\tilde{y}^{\mathbb{N}})$ one has $y_i = \tilde{y}_i$.

2. *intragenerational decision problems* each in the form of a finite feasible set $F \subseteq Y^{\mathbb{N}}$ of consequence profiles having the property that there exist a time $t \in T$, a fixed single generation $\bar{G}_t \subseteq \mathbb{N}$ at time t, and a fixed consequence stream $\bar{y}^{\mathbb{N}} \in Y^{\mathbb{N}}$ such that

$$y^{\mathbb{N}} \in F \Longrightarrow G_t(y^{\mathbb{N}}) = \bar{G}_t \text{ and } y_i = \bar{y}_i \text{ for all } i \notin \bar{G}_t$$

Of course, a one-generation situation is a special case of an intragenerational decision problem. All are decision problems that have a domain of personal consequences for the individuals in $\bar{G}_t$, whereas individuals in all other generations face the fixed personal consequence profile $\langle \bar{y}_i \rangle_{i \in \mathbb{N} \setminus \bar{G}_t}$. For this domain, after ignoring the constant sum $\sum_{i \in \mathbb{N} \setminus \bar{G}_t} u(\bar{y}_i)$, our total utilitarian social welfare function (28) can be reduced to the equivalent *intragenerational* welfare function

$$W_t(y^{\bar{G}_t}) := \sum_{i \in \bar{G}_t} u(y_i) \tag{33}$$

for generation $\bar{G}_t$. This notation, however, does not reflect how the population $G_t(y^{\mathbb{N}})$ of the tth generation depends in principle on the entire personal consequence stream $y^{\mathbb{N}}$. To recognize this dependence, we consider instead the function

$$W_t(y^{\mathbb{N}}) := \sum_{i \in G_t(y^{\mathbb{N}})} u(y_i) \tag{34}$$

By Definition 1 and Assumption 2, the generation $G_t(y^{\mathbb{N}})$ born at each date $t \in T$ is finite. So the function $Y^{\mathbb{N}} \ni y^{\mathbb{N}} \mapsto W_t(y^{\mathbb{N}})$ is well defined on the entire Cartesian product space $Y^{\mathbb{N}}$.

6.4 Intergenerational Welfare

Using the definition (34), the infinite sum (28) of all potential individuals' utilities can be expressed as the sum

$$W_\infty^{\text{total}}(y^{\mathbb{N}}) = \sum_{i=1}^{\infty} u(y_i) = \sum_{t=1}^{\infty} \sum_{i \in G_t(y^{\mathbb{N}})} u(y_i) = \sum_{t=1}^{\infty} W_t(y^{\mathbb{N}}) \tag{35}$$

of the infinite sequence of all generations' welfare levels.

An obvious special case occurs when there is a specified finite *horizon* or *extinction date* H with the property that, for all $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ and all $i \in \cup_{t>H} G_t(y^{\mathbb{N}})$, the only feasible consequence satisfies $y_i = y_0$, thus precluding the possibility that any individual i could ever be born after date H. In effect, this is a restriction on the consequence domain and so on the domain of permissible decision trees.

In this case one has $u(y_i) \equiv 0$ whenever $i \in \cup_{t>H} G_t(y^{\mathbb{N}})$. It follows that, for all permissible personal consequence profiles $y^{\mathbb{N}}$, the total utilitarian objective (28) can be written as the finite sum

$$W_H^{\text{total}}(y^{\mathbb{N}}) = \sum\nolimits_{t=1}^{H} \sum\nolimits_{i \in G_t(y^{\mathbb{N}})} u(y_i) = \sum\nolimits_{t=1}^{H} W_t(y^{\mathbb{N}}) \qquad (36)$$

of the welfare levels for the first H generations, as defined by (34).

6.5 Total Utilitarianism

Faced with the objections to discounting raised by Sidgwick and Ramsey, not to mention the violation of the Suppes equity criterion, it may be worth trying to define intergenerational welfare as the unweighted sum of different generations' welfare, not only with a finite horizon H, as in (36), but also with an infinite horizon. That is, one considers the infinite sum

$$W_\infty^{\text{total}}(y^{\mathbb{N}}) := \sum\nolimits_{i=1}^{\infty} u(y_i) \qquad (37)$$

introduced in (28), even when there is no upper bound n on the number of individuals who may come into existence.

Now, convergence of (37) with infitely many individuals requires that one must have $u(y_i) \to 0$ as $i \to \infty$. This is a version of Parfit's (1984) "repugnant conclusion", which we discussed in Sections 5.2 and 5.3. Nevertheless, one may be able to avoid such repugnance. Indeed, suppose that the personal consequence stream $y^{\mathbb{N}}$ has the property that the finite sum $W_H^{\text{total}}(y^{\mathbb{N}})$ given by (36) reaches a maximum w.r.t. H at a particular extinction date H^*. This suggests that, *ceteris paribus*, even if eventual extinction could somehow be prevented, it would be better to allow it to occur anyway. In other words, extinction would become the only acceptable escape from the repugnant conclusion.

6.6 Ramsey on a Restricted Domain

In (8) of Section 3.8 we introduced the assumption justified by Arrow (1951, 1965, 1971, 1972) that the common utility function $y \mapsto u(y)$ of each individual $i \in \mathbb{N}$ should be bounded both above and below. In particular, there should be an upper bound $\bar{u}$. Inspired by Ramsey (1928) and especially by Gale (1967, p. 11), consider now the restricted domain $\mathcal{D} \subset Y^{\mathbb{N}}$ of "good" personal consequence streams $y^{\mathbb{N}}$ for which, by definition, the expected sum

$$\sum\nolimits_{i \in \mathbb{N}} [u(y_i) - \bar{u}] \tag{38}$$

of non-positive terms is bounded below, and so converges.[20] For finite decision trees with consequence profiles in the domain $\mathcal{D}$, the axioms of Section 3.6, when combined with Suppes equity and the treatment in Section 5 of demographic consequences, together imply that behaviour in any finite decision tree whose consequences are all "good" should maximize the well-defined expected value of (38).

Such "good" consequence streams, however, are likely to be rare, especially in a world where past choices have already committed current generations to future climate change. After all, the series in (38) is bounded below only if the utilities $u(y_i)$ of successive personal consequences y_i to converge to the upper bound $\bar{u}$ as $i \to \infty$. In particular, infinitely many individuals must come into existence, and yet their utilities must converge over time to the upper bound $\bar{u}$. This is surely the antithesis of Parfit's repugnant conclusion.

6.7 Overtaking, Catching Up, and Beyond

Suppose that $\sum_{i=1}^{n} [u(y_i) - \bar{u}]$ diverges to $-\infty$ for all feasible $y^{\mathbb{N}} \in Y^{\mathbb{N}}$, meaning that there are no good consequence streams. In this case the domain $\mathcal{D}$ is empty, so maximizing the expected value of (38) is meaningless. This has led several researchers to suggest alternative concepts of social preference.

First, von Weizsäcker (1965) suggested the *overtaking criterion*, later axiomatized by Brock (1970). Extended to our framework with risky consequences, this states that one extended lottery $\lambda \in \Delta^*(Y^{\mathbb{N}})$ should be strictly preferred to an alternative extended lottery $\mu \in \Delta^*(Y^{\mathbb{N}})$ just in case there exists $h \in \mathbb{N}$ such that, for all $n > h$, the successive differences in total expected values satisfy

$$\sum\nolimits_{y^{\mathbb{N}} \in Y^{\mathbb{N}}} [\lambda(y^{\mathbb{N}}) - \mu(y^{\mathbb{N}})] \sum\nolimits_{i=1}^{n} u(y_i) > 0 \tag{39}$$

To make it compatible with maximizing (38) when there are good consequences, Gale (1967) introduced the weaker *catching up criterion*, according

[20] Ramsey (1928) used a continuous time version of this objective.

to which $\lambda \in \Delta^*(Y^{\mathbb{N}})$ should be weakly preferred to $\mu \in \Delta^*(Y^{\mathbb{N}})$ just in case

$$\liminf_{n\to\infty} \sum\nolimits_{y^{\mathbb{N}}\in Y^{\mathbb{N}}} [\lambda(y^{\mathbb{N}}) - \mu(y^{\mathbb{N}})] \sum\nolimits_{i=1}^{n} u(y_i) \geq 0 \tag{40}$$

Very recently Jonsson and Voorneveld (2018) have introduced *limited discounted utilitarianism* (LDU), based on the mathematical concept of Abel summation. Their LDU criterion states that λ should be weakly preferred to μ just in case, as the discount factor δ tends to 1 from below, the limit infimum of the difference in the respective expected discounted sum of all generations' total welfare levels satisfies

$$\liminf_{\delta\to 1^-} \sum\nolimits_{y^{\mathbb{N}}\in Y^{\mathbb{N}}} [\lambda(y^{\mathbb{N}}) - \mu(y^{\mathbb{N}})] \sum\nolimits_{t=1}^{\infty} \delta^t \sum\nolimits_{i\in G_t(y^{\mathbb{N}})} u(y_i) \geq 0 \tag{41}$$

It has been argued on pragmatic grounds, by Peter Diamond and others, that an appropriate welfare criterion would focus, *ceteris paribus*, on short-run consequences, without worrying unduly about existence or completeness over an infinite horizon. In this spirit, the decade of the 1970s saw several other proposed criteria for planning over an infinite horizon. These include the "agreeable plans" considered by Hammond and Mirrlees (1973) and Hammond (1975), where the loss from not knowing a true finite horizon in advance would tend to zero as the notice of the true horizon tends to infinity.[21] The concept of a "fairly good" plan due to Mirrlees and Stern (1972) was similar; the main difference was in measuring the loss, not as a loss of utility, but as something related to Debreu's (1951) "coefficient of resource utilization" applied to the quantity of the only capital good in their framework. Finally, instead of infinite horizon agreeable plans, Hammond (1973, chs. 9–10; 1974) considered the relative performance of finite horizon "overtures", based on how they could be extended into plans for arbitrarily long finite horizons.

To conclude this section, we remark that none of these procedures provides a satisfactory way of dealing with the "cake eating problem" set out in Gale (1967, p. 4, Example 2). We also recall Chichilnisky's (1997; 2009b, p. 17) important reminder that von Weizsäcker's overtaking criterion does not give a complete preference ordering; nor do any of the other criteria considered in this section. This incompleteness can lead to behaviour in an unfolding decision tree that contradicts the key second and third axioms of Section 3.6, thus leading to unpredictable consequences. In particular, such unpredictability in an entirely deterministic decision tree seems normatively unacceptable.

[21]See also Hammond and Kennan (1979) for related ideas, and Osumi (1986) for a general synthesis that, *inter alia*, embraces results for both discrete and continuous time.

6.8 Discounted Welfare for an Infinite Horizon

The following definition gives, for a suitably restricted domain, a discounted version of the infinite sum (35) in Section 6.4.

Definition 2. *Let* $\beta^T = \langle \beta_t \rangle_{t=1}^{\infty} \in \mathbb{R}_{++}^{\mathbb{N}}$ *denote any infinite sequence of positive* discount factors. *For each* $\beta^T \in \mathbb{R}_{++}^{\mathbb{N}}$, *the* discounted total population *is defined as the discounted sum*

$$N(y^{\mathbb{N}}; \beta^T) := \sum\nolimits_{t=1}^{\infty} \beta_t \# G_t(y^{\mathbb{N}}) \tag{42}$$

of the numbers of individuals in each future generation, which may be infinite. Let

$$\mathcal{D}(\beta^T) := \{y^{\mathbb{N}} \in Y^{\mathbb{N}} \mid N(y^{\mathbb{N}}; \beta^T) < +\infty\} \tag{43}$$

denote the set of personal consequence streams for which the discounted total population is finite. On this domain, define the discounted intergenerational welfare criterion *as the function*

$$\mathcal{D}(\beta^T) \ni y^{\mathbb{N}} \mapsto V(y^{\mathbb{N}}; \beta^T) := \sum\nolimits_{t=1}^{\infty} \beta_t \sum\nolimits_{i \in G_t(y^{\mathbb{N}})} u(y_i) \tag{44}$$

Of course, for (44) to be a true discounted sum, the discount factors β_t should decrease with t. We will not impose this requirement, however.

In the debate between Ramsey on one side and Arrow on the other, both focused on the important special case when each generation consists of a single representative individual — i.e., for all $t \in T$ and $y^{\mathbb{N}} \in Y^{\mathbb{N}}$, one has $\# G_t(y^{\mathbb{N}}) = 1$. Then a necessary and sufficient condition for the infinite wighted sum $\sum_{t=1}^{\infty} \beta_t u(y_t)$ of bounded utilities to converge absolutely for all $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ without restriction is the familiar requirement that $\sum_{t=1}^{\infty} \beta_t < +\infty$.

More generally, however, when for instance the size $\# G_t(y^{\mathbb{N}})$ of successive generations can be steadily growing, the domain $\mathcal{D}(\beta^T)$ on which the discounted objective function $y^{\mathbb{N}} \mapsto V(y^{\mathbb{N}}; \beta^T)$ specified by (44) is well defined depends on the sequence β^T. Indeed:

Proposition 1. *Given any fixed* $\beta^T \in \mathbb{R}_{++}^{\mathbb{N}}$, *the infinite series specified by* (44) *converges absolutely for every* $y^{\mathbb{N}}$ *in the restricted domain* $\mathcal{D}(\beta^T)$ *specified by* (43).

Proof. As stated in (8), Assumption 1 implies the double inequality $\underline{u} \leq u(y) \leq \bar{u}$ for all $y \in Y$, with $u(y_0) = 0$. Define the bound $B := \max\{-\underline{u}, \bar{u}\}$. Then the double inequality implies that for all $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ and all $i \in \mathbb{N}$ one has $|u(y_i)| \leq B$. So provided that $y^{\mathbb{N}} \in \mathcal{D}(\beta^T)$ and so $N(y^{\mathbb{N}}; \beta^T) < +\infty$, summing over all $i \in \mathbb{N}$ while using definition (42) gives

$$\sum\nolimits_{t=1}^{\infty} \beta_t \sum\nolimits_{i \in G_t(y^{\mathbb{N}})} |u(y_i)| \leq \sum\nolimits_{t=1}^{\infty} \beta_t \, \#G_t(y^{\mathbb{N}}) \, B = N(y^{\mathbb{N}}; \beta^T) \, B \quad (45)$$

This confirms absolute convergence. □

6.9 Why Not Discount the Welfare of Future Generations?

In the case when the total population is n, a finite number, the Suppes equity principle that was introduced in Section 4.2 provides one reason for choosing the unweighted sum (28) in Section 5.7. Apart from the informal argument set out in Section 2.3, our main case for choosing the undiscounted objective (37) over the discounted objective (44) is the idea that having an infinite horizon does not seem a good enough reason by itself to abandon the Suppes equity principle. Especially in our setting where even the discounted objective (44) is only defined on the restricted domain $\mathcal{D}(\beta^T)$.

Another argument for favouring the undiscounted sum (37) is based on the fact that the discounted sum (44) violates the usual strict Pareto principle defined below:

Definition 3. *The general infinite horizon von Neumann–Morgenstern welfare objective $Y^{\mathbb{N}} \ni y^{\mathbb{N}} \mapsto V(y^{\mathbb{N}}) \in \mathbb{R}$ satisfies:*

1. *the* strict Pareto criterion *just in case, whenever the two consequence streams $y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}} \in Y^{\mathbb{N}}$ satisfy*
   (a) *$u(y_i) = u(\tilde{y}_i)$ for all $i \in \mathbb{N}$, then $V(y^{\mathbb{N}}) = V(\tilde{y}^{\mathbb{N}})$;*
   (b) *$u(y_i) \geq u(\tilde{y}_i)$ for all $i \in \mathbb{N}$, then $V(y^{\mathbb{N}}) \geq V(\tilde{y}^{\mathbb{N}})$, with $V(y^{\mathbb{N}}) > V(\tilde{y}^{\mathbb{N}})$ except when $u(y_i) = u(\tilde{y}_i)$ for all $i \in \mathbb{N}$;*
2. *the* restricted Pareto criterion *just in case it satisfies the strict Pareto criterion whenever the two consequence streams $y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}} \in Y^{\mathbb{N}}$ are restricted to satisfy $G_t(y^{\mathbb{N}}) = G_t(\tilde{y}^{\mathbb{N}})$ for all $t \in T$.*

The reason why (44) violates the strict Pareto principle is that the weight attached to the individual who is given the number i is β_t, for the unique date t such that $i \in G_t(y^{\mathbb{N}})$. Of course, this t is the date at which the ith individual is born when the personal consequence sequence is $y^{\mathbb{N}}$. Typically, of course, the relative weights of any two numbered individuals $i, j \in \mathbb{N}$ therefore depend on $y^{\mathbb{N}}$. This reflects the failure to give each numbered individual $i \in \mathbb{N}$ a proper identity until one specifies the generation $G_t(y^{\mathbb{N}})$ to which i belongs when the personal consequence stream is $y^{\mathbb{N}}$.

By contrast, the restricted Pareto condition defined in the second part of Definition 3 applies only when each individual's generation is not affected by the choice between $y^{\mathbb{N}}$ and $\tilde{y}^{\mathbb{N}}$. Of course, in the special case when each numbered individual $i \in \mathbb{N}$ has a fixed date of birth, and so $G_t(y^{\mathbb{N}}) = G_t$, independent of $y^{\mathbb{N}}$, then the restricted Pareto criterion coincides with the strict Pareto criterion. Indeed, in this special case, the restricted domain (43) becomes

$$\mathcal{D}(G^T) := \{y^{\mathbb{N}} \in Y^{\mathbb{N}} \mid \sum_{t=1}^{\infty} \beta_t \# G_t < +\infty\} \tag{46}$$

which depends only on the constant generational structure. On this domain, the discounted sum (44) takes the form

$$V(y^{\mathbb{N}}; \beta^T) = \sum_{t=1}^{\infty} \beta_t \sum_{i \in G_t} u(y_i) \tag{47}$$

We remark that this is a biased original position of the kind discussed in Section 3.5, with the probability of becoming any particular individual $i \in G_t$ given by

$$\mu_i = \beta_t / \sum_{s=1}^{\infty} \beta_s \, \# G_s \tag{48}$$

7 Sustainability

7.1 Sustainable Development: Motivation

There is by now an overwhelming scientific consensus that the world risks being engulfed by climate change — see Stern (2006, 2008, 2013, 2015) for a review of the relevant literature. Even humanity's future viability could be at stake. Recognizing this threat takes us well beyond what had previously been the usual realm of social choice or economic theory. Indeed, the problem pits increases in present consumption against increases in the likelihood of future existence. This brings a new sense of urgency to Ramsey's famous ethical objection to discounting, anticipated somewhat by Sidgwick and Pigou.

This has propelled the nations of the world to contemplate the practical implications of the theoretical issue regarding discounting that Ramsey, Suppes, Koopmans and Arrow, amongst others, disputed some time ago. In Brundtland (1987), sustainable development was famously defined as:

> ... development that meets the needs of the present without compromising the ability of future generations to meet their own needs.

Later, after the UN General Assembly had asked for a report on the progress toward sustainable development, in June 1992 the United Nations Conference on Environment and Development (UNCED, or the "Earth Summit") was convened in Rio De Janeiro. That was where 162 nations adopted and signed the United Nations Framework Convention on Climate Change (UNFCCC). That same year saw the creation of the Group of 20, consisting of the 19 nations with the world's largest economies, plus the European Union at a time when it consisted of 12 member states. In the charter of the new G20, Sustainable Development was identified as the top priority of international economic development. Subsequently, in the year 2015 when the period covered by its Millennium Development Goals was expiring, the United Nations held a Sustainable Development Summit at which these goals were replaced with a new document entitled "Transforming our world: the 2030 Agenda for Sustainable Development".

7.2 Sustainable Development and Maximin

Decades before Brundtland (1987), Hicks (1946, p. 174) had an idea similar to sustainability when he defined an individual's "income" as

> ... the maximum amount of money which the individual can spend this week, and still expect to be able to spend the same amount in real terms in each ensuing week.

In this spirit, and following Solow (1991, 2012), sustainability might be defined as giving each generation access to an opportunity set that allows it to be no worse off than it would have been with the opportunity set that was available to any of its predecessors. This suggests trying to maximize the initial generation's welfare level subject to *monotone sustainability* — i.e., requiring successive generations' welfare levels to be non-decreasing over time.[22]

In many settings an efficient development path turns out to be monotone sustainable if and only if it satisfies the "Rawlsian maximin" criterion of maximizing the minimum welfare level over all future generations. This criterion for optimal saving was first used in the independent contributions

[22]For a fuller discussion of sustainability in this sense, see *inter alia* Hammond (1993) as well as Pezzey (1997, p. 451) and Asheim (2007).

of Arrow (1973b) and, in the continuous time case, of Solow (1974). In many cases, the maximin path will be “regular” in the sense that all successive generations have the same utility, as in Burmeister and Hammond (1977). For an interesting special case, any regular maximin path in continuous time can be characterized by “Hartwick’s rule” for resource depletion, requiring all rents from exhaustible resource to be invested in new capital stock — see Hartwick (1977) as well as Dixit, Hammond and Hoel (1980) and, for extensive and enlightening discussions of some basic issues, Heal (1998) and Asheim (2007). More recently, Llavador, Roemer and Silvestre (2010, 2011, 2015) have made maximin utility a key feature of their discussion of optimal policies in the face of climate change.

7.3 Interpersonal Leximin

The growth literature on sustainability described above focused on generations. Then, however, a regular maximin path on which each generation has the same welfare level is consistent with some members of each generation facing very bad personal consequences. Indeed, while the phrase from paragraph 1 of chapter 2 of the Brundtland Report that was reproduced in Section 7.1 has often been quoted, much less attention has been devoted to this sentence from paragraph 4:

> Sustainable development requires meeting the basic needs of all and extending to all the opportunity to satisfy their aspirations for a better life.[23]

Thus, especially in this paper on social choice, rather than the *intergenerational* maximin criterion that maximizes the infimum

$$\inf_{t \in T} w_t \left(\langle y_i \rangle_{i \in G_t(y^{\mathbb{N}})} \right) \tag{49}$$

over all generations’ welfare levels, instead it seems both better and simpler to apply the *interpersonal* maximin criterion that maximizes the infimum

$$\inf_{i \in \mathbb{N}} u(y_i) \tag{50}$$

over all persons’ utility levels. We discuss only the maximand (50) from now on. Moreover, we consider a social welfare ordering $\succsim$ defined on the domain

[23] This calls to mind the Basic Needs Model of the Bariloche Foundation in Argentina, which had been developed by an interdisciplinary team led by Amilcar Herrera that also included GC. See Herrera *et al.* (1976) and Chichilnisky (1977).

$[\underline{u}, \bar{u}]^{\mathbb{N}}$ of utility sequences $u^{\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ that are uniformly bounded below by $\underline{u}$ and above by $\bar{u}$.

Recall that, following Arrow (1963), social choice theory typically uses the weak Pareto condition

$$u^{\mathbb{N}} \gg \tilde{u}^{\mathbb{N}} \Longrightarrow u^{\mathbb{N}} \succ \tilde{u}^{\mathbb{N}} \tag{51}$$

rather than the strict Pareto condition

$$u^{\mathbb{N}} > \tilde{u}^{\mathbb{N}} \Longrightarrow u^{\mathbb{N}} \succ \tilde{u}^{\mathbb{N}} \tag{52}$$

which Diamond (1965) called "sensitivity".[24] Evidently maximin satisfies the weak condition (51) for a finite set of individuals. It need not, however, with an infinite set of individuals because the strict inequality $u^{\mathbb{N}} \gg \tilde{u}^{\mathbb{N}}$ in the antecedent of (51) can hold even when $\inf_i \tilde{u}_i = \inf_i u_i$.

In order to strengthen the maximin criterion to satisfy the strict Pareto condition (52) in a finite population, Rawls (1971) himself suggested that one should consider "leximin" — i.e., the lexicographic extension of maximin. In a finite population of m individuals, for each *rank* $r \in \mathbb{N}_m$ we can define the rth lowest utility level $\ell_r(u^{\mathbb{N}_m})$, breaking ties arbitrarily. Then the *lexicographic strict ordering* $>_L$ on the m-dimensional space $\mathbb{R}^m$ is defined so that

$$\begin{aligned} u^{\mathbb{N}_m} >_L \tilde{u}^{\mathbb{N}_m} \Longleftrightarrow \exists k \in \mathbb{N}_m : \ell_r(u^{\mathbb{N}_m}) &= \ell_r(\tilde{u}^{\mathbb{N}_m}) \\ \text{for } r = 1, 2, \ldots k-1 \quad \text{and} \quad & \ell_k(u^{\mathbb{N}_m}) > \ell_k(\tilde{u}^{\mathbb{N}_m}) \end{aligned} \tag{53}$$

Finally, the *leximin* rule requires maximizing the complete and transitive weak preference ordering defined by

$$u^{\mathbb{N}_m} \succsim \tilde{u}^{\mathbb{N}_m} \Longleftrightarrow \neg[\tilde{u}^{\mathbb{N}_m} >_L u^{\mathbb{N}_m}] \tag{54}$$

A fundamental difficulty in extending the leximin rule to infinite utility streams in $\mathbb{R}^{\mathbb{N}}$ comes in defining appropriate ranks on an infinite sequence, especially one that has no minimum element. These issues have been carefully discussed by, amongst others, Asheim (2007, 2010), Asheim, Mitra and Tungodden (2012), Zuber and Asheim (2012), as well as Asheim and Zuber (2013).

[24]We adopt the notational convention that $u^{\mathbb{N}} \gg \tilde{u}^{\mathbb{N}} \Longleftrightarrow u_i > \tilde{u}_i$ for all $i \in \mathbb{N}$, whereas $u^{\mathbb{N}} > \tilde{u}^{\mathbb{N}} \Longleftrightarrow u_i \geq \tilde{u}_i$ for all $i \in \mathbb{N}$, with $u^{\mathbb{N}} \neq \tilde{u}^{\mathbb{N}}$.

7.4 Sustainable Welfare Criteria

Rather than sustainable plans that maximize a particular welfare criterion like leximin, Chichilnisky (1996) pioneered the study of sustainable welfare criteria, introduceing two axioms which require that neither the present nor the future should play a dictatorial role. GC also characterized all the "sustainable" welfare criteria that these axioms imply. Such preferences exist, are readily computable, and have several other desirable properties.

In the framework of this paper, it is natural to consider sustainable preferences over lotteries with infinite-horizon sequences $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ of personal consequences as outcomes. Moreover, the relevant welfare criterion should be the expected value of some uniformly bounded von Neumann–Morgenstern social utility function

$$Y^{\mathbb{N}} \supseteq \mathcal{D} \ni y^{\mathbb{N}} \mapsto V(y^{\mathbb{N}}) \in \mathbb{R} \tag{55}$$

defined on a possibly restricted domain $\mathcal{D}$ of personal consequence streams $y^{\mathbb{N}}$ that belong to the countably infinite Cartesian product set $Y^{\mathbb{N}}$. Moreover, the function $y^{\mathbb{N}} \mapsto V(y^{\mathbb{N}})$ should satisfy the restricted Pareto criterion set out in part 2 of definition 3.

7.5 Definition of Temporal Dictatorship

Following Chichilnisky (1996, pp. 240–241), a "dictatorship of the present" is a welfare criterion which, after some generation that depends on the choices at hand, is insensitive to the welfare of all succeeding generations. In other words, a dictatorship of the present occurs if a strict preference for one personal consequence stream $y^{\mathbb{N}}$ over an alternative stream $\tilde{y}^{\mathbb{N}}$ cannot be overturned by any changes in these two streams that affect only sufficiently distant generations.

By contrast, a "dictatorship of the future" is insensitive to the welfare of the present, disregarding the welfare of all generations that precede some generation. In other words, a dictatorship of the future occurs if a strict preference for one personal consequence stream $y^{\mathbb{N}}$ over an alternative stream $\tilde{y}^{\mathbb{N}}$ cannot be overturned by any changes in these two streams that affect only generations that are sufficiently close to the present.

Using alternative terminology suggested by Heal (1998, p. 69), a criterion displaying dictatorship of the present is *insensitive to the long-run future*; whereas one displaying dictatorship of the future is *insensitive to the present*. We claim that neither form of temporal dictatorship is ethically acceptable. Somewhat surprisingly, it is relatively easy to find welfare criteria that are sustainable in the sense that both forms of temporal dictatorship are avoided.

To complete this section, we adapt Chichilnisky's original formal definitions of both forms of temporal dictatorship to the setting here, where the generational structure G^T depends on the personal consequence stream $y^{\mathbb{N}}$.

Definition 4. *A welfare criterion $V : \mathcal{D} \to \mathbb{R}$ on the domain $\mathcal{D} \subseteq Y^{\mathbb{N}}$ is a* dictatorship of the present *if for all pairs $y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}} \in \mathcal{D}$ with $V(y^{\mathbb{N}}) > V(\tilde{y}^{\mathbb{N}})$, there exists a date $s = s(y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}}) \in T$ such that whenever the pair $z^{\mathbb{N}}, \tilde{z}^{\mathbb{N}} \in Y^{\mathbb{N}}$ satisfies*

$$i \in \cup_{t=1}^{s} G_t(y^{\mathbb{N}}) \Longrightarrow z_i = y_i \quad \text{and} \quad i \in \cup_{t=1}^{s} G_t(\tilde{y}^{\mathbb{N}}) \Longrightarrow \tilde{z}_i = \tilde{y}_i \tag{56}$$

then $z^{\mathbb{N}}, \tilde{z}^{\mathbb{N}} \in \mathcal{D}$ and $V(z^{\mathbb{N}}) > V(\tilde{z}^{\mathbb{N}})$.

Definition 5. *A welfare criterion $V : \mathcal{D} \to \mathbb{R}$ on the domain $\mathcal{D} \subseteq Y^{\mathbb{N}}$ is a* dictatorship of the future *if for all pairs $y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}} \in \mathcal{D}$ with $V(y^{\mathbb{N}}) > V(\tilde{y}^{\mathbb{N}})$, there exists a date $s = s(y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}}) \in \mathbb{N}$ such that whenever the pair $z^{\mathbb{N}}, \tilde{z}^{\mathbb{N}} \in Y^{\mathbb{N}}$ satisfies*

$$i \in \cup_{t=s}^{\infty} G_t(y^{\mathbb{N}}) \Longrightarrow z_i = y_i \quad \text{and} \quad i \in \cup_{t=s}^{\infty} G_t(\tilde{y}^{\mathbb{N}}) \Longrightarrow \tilde{z}_i = \tilde{y}_i \tag{57}$$

then $z^{\mathbb{N}}, \tilde{z}^{\mathbb{N}} \in \mathcal{D}$ and $V(z^{\mathbb{N}}) > V(\tilde{z}^{\mathbb{N}})$.

Note that the only difference between these two definitions concerns the conditions (56) and (57). These determine whether the strict preference between two personal consequence streams remains unchanged after alterations in the consequence streams only for all generations that originate either: (i) after date s, so sufficiently far into the future; or (ii) before date s, so sufficiently close to the present.

7.6 Two Examples of Temporal Dictatorship

Our first example extends to our framework the argument in Chichilnisky (1996) that discounting the utilities of future generations leads to a dictatorship of the present.

Example 1. *For any sequence of discount factors $\beta^T \in \mathbb{R}^T_{++}$, consider the restricted domain $\mathcal{D}(\beta^T)$ specified by* (43) *in Section 6.8, and the infinite-horizon discounted welfare criterion $\mathcal{D}(\beta^T) \ni y^{\mathbb{N}} \mapsto V(y^{\mathbb{N}}; \beta^T) \in \mathbb{R}$, which is defined by* (44)*. To show that this criterion is a dictatorship of the present, let $y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}} \in \mathcal{D}(\beta^T)$ be any two personal consequence streams which satisfy*

$V(y^{\mathbb{N}};\beta^T) - V(\tilde{y}^{\mathbb{N}};\beta^T) =: \delta > 0$. By definition (43)*, because $y^{\mathbb{N}}, \tilde{y}^{\mathbb{N}} \in \mathcal{D}(\beta^T)$, one can choose $s \in \mathbb{N}$ sufficiently large to ensure that*

$$\sum\nolimits_{t=s}^{\infty} \beta_t \left[\#G_t(y^{\mathbb{N}}) + \#G_t(\tilde{y}^{\mathbb{N}})\right] < \tfrac{1}{2}\delta/(\bar{u} - \underline{u}) \tag{58}$$

Consider next two more personal consequence streams $z^{\mathbb{N}}, \tilde{z}^{\mathbb{N}} \in \mathcal{D}(\beta^T)$ that satisfy (56) *for this particular value of $s \in \mathbb{N}$. By definition* (44)*, one has*

$$\begin{aligned} & V(z^{\mathbb{N}};\beta^T) - V(\tilde{z}^{\mathbb{N}};\beta^T) \\ & \qquad = \sum\nolimits_{t=1}^{\infty} \beta_t \left[\sum\nolimits_{i \in G_t(z^{\mathbb{N}})} u(z_i) - \sum\nolimits_{i \in G_t(\tilde{z}^{\mathbb{N}})} u(\tilde{z}_i)\right] \end{aligned} \tag{59}$$

By Assumption 2 in Section 6.2, for all $t \leq s$ it follows from (56) *that $G_t(z^{\mathbb{N}}) = G_t(y^{\mathbb{N}})$ and $G_t(\tilde{z}^{\mathbb{N}}) = G_t(\tilde{y}^{\mathbb{N}})$ and also*

$$V(z^{\mathbb{N}};\beta^T) - V(\tilde{z}^{\mathbb{N}};\beta^T) = V(y^{\mathbb{N}};\beta^T) - V(\tilde{y}^{\mathbb{N}};\beta^T) + \Sigma_s = \delta + \Sigma_s \tag{60}$$

where Σ_s is defined as the discounted sum

$$\sum\nolimits_{t=s}^{\infty} \beta_t \left[\sum\nolimits_{i \in G_t(y^{\mathbb{N}})} [u(z_i) - u(y_i)] - \sum\nolimits_{i \in G_t(\tilde{y}^{\mathbb{N}})} [u(\tilde{z}_i) - u(\tilde{y}_i)]\right] \tag{61}$$

of differences between: (i) the total welfare change of all individuals from generation s on when consequence stream $y^{\mathbb{N}}$ is replaced by $z^{\mathbb{N}}$; (ii) the corresponding total welfare change when consequence stream $\tilde{y}^{\mathbb{N}}$ is replaced by $\tilde{z}^{\mathbb{N}}$.

Now the boundedness assumption (8) *implies that for all $y, y' \in Y$ one has both $u(y) \geq \underline{u}$ and $-u(y') \geq -\bar{u}$, so*

$$u(y) - u(y') \geq \underline{u} - \bar{u} = -(\bar{u} - \underline{u}) \tag{62}$$

It follows from (62) *and definition* (61) *that*

$$\Sigma_s \geq -\sum\nolimits_{t=s}^{\infty} \beta_t \left[\#G_t(y^{\mathbb{N}}) + \#G_t(\tilde{y}^{\mathbb{N}})\right] (\bar{u} - \underline{u}) \tag{63}$$

So choosing s to satisfy (58) *implies that $\Sigma_s > -\frac{1}{2}\delta$. Then* (60) *implies that*

$$V(z^{\mathbb{N}};\beta^T) - V(\tilde{z}^{\mathbb{N}};\beta^T) > \delta - \tfrac{1}{2}\delta = \tfrac{1}{2}\delta > 0 \tag{64}$$

This confirms that there is a dictatorship of the present.

Our second example is a simple adaptation of the result due to Chichilnisky (1997) concerning dictatorship of the future.

Example 2. *Because of the boundedness assumption* (8)*, the two alternative welfare criteria*

$$V_*(y^{\mathbb{N}}) := \liminf_{i\to\infty} u(y_i) \quad \text{and} \quad V^*(y^{\mathbb{N}}) := \limsup_{i\to\infty} u(y_i) \tag{65}$$

are both well defined on the unrestricted domain $Y^{\mathbb{N}}$ *of all personal consequence streams. Indeed, for all* $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ *one has*

$$\underline{u} \leq V_*(y^{\mathbb{N}}) \leq V^*(y^{\mathbb{N}}) \leq \bar{u} \tag{66}$$

Also, given any $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ *and any* $s \in \mathbb{N}$*, consider the set* $\cup_{t=1}^{s} G_t(y^{\mathbb{N}})$ *of all individuals who, given the personal consequence sequence* $y^{\mathbb{N}}$*, make up the first* s *generations. Then it is obvious that, for all* i *in this set, both values* $V_*(y^{\mathbb{N}})$ *and* $V^*(y^{\mathbb{N}})$ *are independent of* y_i*. Hence the two welfare criteria* $Y^{\mathbb{N}} \ni y^{\mathbb{N}} \mapsto V_*(y^{\mathbb{N}})$ *and* $Y^{\mathbb{N}} \ni y^{\mathbb{N}} \mapsto V^*(y^{\mathbb{N}})$ *both display dictatorship of the future.*

7.7 Sustainable von Neumann–Morgenstern Social Utility

Chichilnisky's (1996) definition of sustainable preferences applies when:

1. consequences and their utilities are determinate;
2. the generational structure is equivalent to one where there is one representative individual for each generation;
3. the relevant welfare criterion is a (complete and transitive) preference ordering over the domain of all possible utility streams which is also "sensitive", in the sense of satisfying the strict Pareto criterion.

The following definition adapts that of Chichilnisky (1996) to the current setting where the social objective is the expected value of a von Neumann–Morgenstern social utility function which is defined on a restricted domain $\mathcal{D} \subset Y^{\mathbb{N}}$, and where each admissible *ex post* personal consequence stream $y^{\mathbb{N}} \in \mathcal{D}$ determines a corresponding generation structure $G^T(y^{\mathbb{N}})$.

Definition 6. *Given a restricted domain* $\mathcal{D} \subset Y^{\mathbb{N}}$ *of admissible personal consequence streams, a* sustainable *von Neumann–Morgenstern social utility function is a mapping* $\mathcal{D} \ni y^{\mathbb{N}} \mapsto V(y^{\mathbb{N}}) \in \mathbb{R}$ *which:*

1. *satisfies the* restricted Pareto criterion *set out in part 2 of definition 3;*
2. *is neither a dictatorship of the present, nor a dictatorship of the future.*

The following example offers a two-dimensional parametric class of sustainable preferences.

Example 3. *For each stream of discount factors $\beta^T \in \mathbb{R}^T_{++}$, consider the restricted domain $\mathcal{D}(\beta^T)$ specified by (43) of personal consequence streams $y^\mathbb{N} \in Y^\mathbb{N}$ for which the total discounted population is finite. Then, for any pair of parameters $\alpha \in [0,1]$ and $\omega \in (0,1)$, consider the von Neumann–Morgenstern welfare function*

$$\mathcal{D}(\beta^T) \ni y^\mathbb{N} \mapsto \Psi(y^\mathbb{N}; \beta^T, \alpha, \omega) := (1-\omega) \sum\nolimits_{t \in T} \beta_t \sum\nolimits_{i \in G_t(y^\mathbb{N})} u(y_i) + \omega \left[\alpha \liminf_{i \to \infty} u(y_i) + (1-\alpha) \limsup_{i \to \infty} u(y_i) \right] \quad (67)$$

This puts positive weight $1-\omega$ on the first term, which is a dictatorship of the present, and positive weight ω on the second term, which is a dictatorship of the future. The strict convex combination of these two defines sustainable preferences. The welfare function satisfies the restricted Pareto criterion introduced in Definition 3 of Section 6.9.

Note that the second term of (67) is an asymptotic form of the "Hurwicz criterion" for decisions under uncertainty that was discussed, *inter alia*, in Arrow and Hurwicz (1972). Here the parameter α can be regarded as a "coefficient of pessimism".

Going beyond Example 3, Chichilnisky (1996) characterizes sustainable preferences for the important special case when the preference ordering over sure consequence streams is represented by a utilitarian social welfare function defined on the Banach space ℓ^∞ of all bounded utility streams in $\mathbb{R}^\mathbb{N}$ that is not only strictly increasing and continuous, but is also linear. For this important special case, Theorem 2 of Chichilnisky (1996) offers a complete characterization using finitely additive measures that represent linear functionals in the dual of ℓ^∞. See Chichilnisky (1996, 1997, 2009), Heal (2000), and Lauwers (2017) for further discussion, including specific examples that use sustainable preferences to analyse renewable and exhaustible resources.

Among other work that discusses sustainable preferences, we mention Figuières and Tidball (2012), which builds on work by Chichilnisky, Heal and Beltratti (1995) that considers "the green golden rule".

8 Extinction Discounting and Beyond

8.1 The Idea of Extinction Discounting

The main approach to discounting the welfare of future generations that is advocated in the *Stern Review* on the economics of climate change can be summarized in the following three passages:

> ... while we do allow, for example, for the possibility that, say, a meteorite might obliterate the world, and for the possibility that future generations may be richer (or poorer), we treat the *welfare* of future generations on a par with our own." Stern (2006, p. 35)

> ... the only sound ethical basis for placing less value on the utility (as opposed to consumption) of future generations was the uncertainty over whether or not the world will exist, or whether those generations will all be present. Stern (2006, p. 51)

> Where discount rates are used in modeling the economic benefits of climate policy, they should use consumption as the numeraire and adopt a pure-time discount rate close to zero (the small positive value reflecting the risk of planetary annihilation only). Stern (2015, p. 174)

The basic idea for the rule that one should discount only to reflect the probability of extinction can be traced back to an essay by Ramsey (1931, p. 291), as indicated in the following passage from Arrow (1999b):[25]

> Koopmans, who has in fact given the basic argument *for* discounting, nevertheless holds "an ethical preference for neutrality as between the welfare of different generations" (1965, p. 239). Robert Solow (1974, p. 9) [wrote]: "In solemn conclave assembled, so to speak, we ought to act as if the social rate of time preference were zero." When the conclave is not so solemn, different thoughts appear. Ramsey presented a talk to a group of friends at Cambridge (the Society, frequently referred to as the Apostles), in which, talking about our observations of the universe, he said: "My picture of the world is drawn in perspective. ... I apply my perspective not merely to space but also to time. In time the world will cool and everything will die; but that is a long way off still, and its present value at compound interest is almost nothing."

[25] Arrow adds emphasis to the end of his quotation from Ramsey, which we have removed.

Fifty years ago Mirrlees (1967, p. 96),[26] in setting out what he regarded as some essential departures from the "famous paper" of Ramsey (1928), provided what seems to be the first clear statement of what we are going to call the "extinction" discounting rule:

> I admit the possibility of discounting at a rate r, to allow for the likelihood of extinction.

The first person, however, to write about the general idea of discounting due to possible non-existence may be Sidgwick (1907, p. 414), who after the passage on the interests of posterity which we quoted in Section 5.6, goes on to add:

> ...except in so far as the effect of his actions on posterity and even the existence of human beings to be affected must necessarily be more uncertain.

8.2 Restricting the Domain of Consequence Lotteries

Once we recognize the unavoidable risk of eventual extinction, it is natural to consider lotteries $\lambda \in \Delta^*(Y^{\mathbb{N}})$ over infinite consequence streams. For each such lottery, the expected value of the unweighted total utilitarian objective (37) is

$$W(\lambda) := \mathbb{E}_\lambda \left[\sum\nolimits_{i=1}^{\infty} u(y_i) \right] = \sum\nolimits_{i=1}^{\infty} \mathbb{E}_\lambda[u(y_i)] \tag{68}$$

provided that the infinite sum is defined. The idea of extinction discounting will be to ensure that for each λ in a suitably restricted domain, and for each sufficient large $i \in \mathbb{N}$, the probability that $y_i = y_0$ because i exceeds the total population before extinction should be high enough to ensure that the infinite series in (68) converges absolutely.

[26]It should be noted that both KJA and Koopmans are recognized in Mirrlees' paper as being "[a]mong many to whom acknowledgement is due for comment and discussion". It also seems appropriate to mention that PJH and NHS both owe an enormous debt to Jim Mirrlees for arousing our interest in optimal growth theory, and for supervising our Ph.D. theses. Finally, we take the opportunity to correct the impression given in some obituaries of Arrow that Mirrlees should be added to Harsanyi, Maskin, Myerson and Spence — the four of KJA's former Ph.D. students who were awarded the Nobel Memorial Prize in Economics during his lifetime. In fact, instead of advising Mirrlees' Ph.D. thesis, the University of Cambridge appointed KJA as external examiner. At the "Academic Tribute to Kenneth Arrow" held at the Stanford Institute of Economic Policy Research on 9th October 2017, Mirrlees duly acknowledged receiving KJA's extensive comments.

8.3 A Finite Expected Population as a Sufficient Condition

For each $\lambda \in \Delta^*(Y^{\mathbb{N}})$, recall the notation λ_i introduced in (3) of Section 3.5 for the marginal distribution that λ induces on the ith component y_i of the personal consequence stream $Y^{\mathbb{N}}$. Recall too that in Section 3.3 we defined y_0 as the unique personal consequence in Y where the person never exists. Hence, the probability that individual i ever joins the population of individuals who eventually exist is $\lambda_i(Y \setminus \{y_0\})$, the probability that $y_i \neq y_0$. Therefore, the expected total number of individuals who do eventually come into existence is

$$N(\lambda) := \sum\nolimits_{i \in \mathbb{N}} \lambda_i(Y \setminus \{y_0\}) \tag{69}$$

This sum, of course, may be infinite. Restricting attention, however, to lotteries $\lambda \in \Delta^*(Y^{\mathbb{N}})$ for which $N(\lambda)$ is finite has the following important implication.

Proposition 2. *Under the boundedness Assumption 1, for each lottery λ such that the expected population $N(\lambda)$ defined by* (69) *is finite, the infinite series $\sum_{i=1}^{\infty} \mathbb{E}_\lambda u(y_i)$ in* (68) *is absolutely convergent.*

Proof. As stated in (8), Assumption 1 implies that $\underline{u} \leq u(y) \leq \bar{u}$ for all $y \in Y$, with $u(y_0) = 0$. Using the definition of marginal probability, it follows that for all $i \in \mathbb{N}$ and all $\lambda \in \mathcal{D}_{\text{fep}}$ one has

$$\lambda_i(Y \setminus \{y_0\})\, \underline{u} \leq \mathbb{E}_{\lambda_i} u(y_i) = \mathbb{E}_\lambda u(y_i) \leq \lambda_i(Y \setminus \{y_0\})\, \bar{u} \tag{70}$$

Define $B := \max\{-\underline{u}, \bar{u}\}$. Then for all $i \in \mathbb{N}$ the double inequality (70) implies that

$$|\mathbb{E}_\lambda u(y_i)| \leq B\, \lambda_i(Y \setminus \{y_0\}) \tag{71}$$

By definition (69), summing (71) over all $i \in \mathbb{N}$ gives

$$\sum\nolimits_{i \in \mathbb{N}} |\mathbb{E}_\lambda u(y_i)| \leq B \sum\nolimits_{i \in \mathbb{N}} \lambda_i(Y \setminus \{y_0\}) = B\, N(\lambda) \tag{72}$$

Provided that $N(\lambda) < +\infty$, this confirms absolute convergence in (68). $\square$

8.4 Stochastic Processes of Extinction

Our argument for extinction discounting relies on treating extinction as a stochastic process. So, for each time $t \in T$, given that extinction will not have occurred already before time t, let η_t denote the conditional probability that extinction will occur precisely at time t. This is the *hazard rate* associated with the extinction process.

Of course, this definition implies that $1 - \eta_t$ is the conditional probability at time $t-1$ of survival for at least one more period up to time t. Then for each time $t \in T$, the multiplicative property of successive conditional probabilities of surviving one more period implies that the probability of survival till at least time t — i.e., the unconditional probability at time $t = 0$ that a non-empty generation G_t can come into existence at time t — is given by

$$\pi_t := \prod_{s=1}^{t} (1 - \eta_s) \tag{73}$$

But then the unconditional probability that extinction occurs precisely at any specific date $t \in T$ is

$$\pi_{t-1} - \pi_t = [1 - (1 - \eta_t)] \prod_{s=1}^{t-1} (1 - \eta_s) = \eta_t \prod_{s=1}^{t-1} (1 - \eta_s) = \pi_t \, \eta_t \tag{74}$$

where $\pi_0 = 1$.

The following definition recognizes how the astrophysics of the sun, together with possible unavoidable asteroid impacts and also cataclysmic supervolcanos, can be assumed to guarantee humanity's ultimate extinction.

Definition 7. *An* extinction process *is a non-increasing sequence* $\pi^T = \langle \pi_t \rangle_{t \in T}$ *of unconditional* survival probabilities $\pi_t \in [0, 1]$ *satisfying* $\pi_t \to 0$ *as* $t \to \infty$*. Let* Π *denote the set of all such processes.*

By (74), the probability of extinction at or before any given time $s \in T$ is

$$\sum_{t=1}^{s} \pi_t \, \eta_t = \sum_{t=1}^{s} (\pi_{t-1} - \pi_t) = \pi_0 - \pi_s = 1 - \pi_s \tag{75}$$

When applied to (75), Definition 7 implies that the probability of ultimate extinction is the limit

$$\sum_{t=1}^{\infty} \pi_t \, \eta_t = \lim_{s \to \infty} \sum_{t=1}^{s} \pi_t \, \eta_t = \lim_{s \to \infty} (1 - \pi_s) = 1 \tag{76}$$

8.5 Extinction Discounting: Basic Definitions

Consider now the the product space $\Pi \times Y^{\mathbb{N}}$ whose members $(\pi^T, y^{\mathbb{N}})$ combine an extinction process with a personal consequence stream. Here each $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ should be interpreted, not as an *actual* infinite stream of lifetime personal consequences, but rather as the unique *potential* stream that would come about only in the zero probability event that extinction is deferred indefinitely.

For each extinction date or *finite horizon* $t \in T$ and each potential consequence stream $y^{\mathbb{N}} \in Y^{\mathbb{N}}$, define the *finite horizon* consequence stream $y^{\mathbb{N}}|_t = \langle y_i|_t \rangle_{i \in \mathbb{N}} \in Y^{\mathbb{N}}$ whose respective personal components are given by

$$y_i|_t := \begin{cases} y_i & \text{if } i \in \cup_{s=1}^t G_s(y^{\mathbb{N}}) \\ y_0 & \text{if } i \notin \cup_{s=1}^t G_s(y^{\mathbb{N}}) \end{cases} \tag{77}$$

In other words, $y^{\mathbb{N}}|_t$ is the modification of $y^{\mathbb{N}}$ that would result if extinction were to occur precisely at date t, thus preventing members of later generations from ever existing.

For each $t \in T$, let $\delta_{y^{\mathbb{N}}|_t} \in \Delta(Y^{\mathbb{N}})$ denote the degenerate lottery that attaches probability 1 to the corresponding finite horizon consequence stream $y^{\mathbb{N}}|_t$. With this notation, depending on $(\pi^T, y^{\mathbb{N}}) \in \Pi \times Y^{\mathbb{N}}$, the particular consequence lottery in $\Delta^*(Y^{\mathbb{N}})$ that eventually results is

$$\lambda(\pi^T, y^{\mathbb{N}}) = \sum\nolimits_{t \in T} \pi_t \, \eta_t \, \delta_{y^{\mathbb{N}}|_t} = \sum\nolimits_{t \in T} (\pi_{t-1} - \pi_t) \, \delta_{y^{\mathbb{N}}|_t} \tag{78}$$

where $\pi_0 = 1$. Because (76) implies that $\sum_{t \in T} \pi_t \, \eta_t = 1$, this is an infinite probability mixture of the list $\langle \delta_{y^{\mathbb{N}}|_t} \rangle_{t \in T}$ of successive degenerate finite-horizon lotteries.

Consider any fixed extinction process $\pi^T \in \Pi$ and consequence stream $y^{\mathbb{N}} \in Y^{\mathbb{N}}$. For any $i \in G_t(y^{\mathbb{N}})$, rather than the certainty of y_i, instead each person $i \in G_t(y^{\mathbb{N}})$ actually faces a lottery given by the probability mixture

$$\lambda_i(\pi^T, y^{\mathbb{N}}) = \pi_t \, \delta_{y_i} + (1 - \pi_t) \, \delta_{y_0} \tag{79}$$

Denote the associated probability of the event $y_i \in Y \setminus \{y_0\}$ that any individual $i \in G_t(y^{\mathbb{N}})$ will finally exist by

$$\sigma_i(\pi^T, y^{\mathbb{N}}) := \pi_t = \lambda_i(\pi^T, y^{\mathbb{N}})(Y \setminus \{y_0\}) \tag{80}$$

Given (80), the *expected population before extinction*, which was given by (69), can also be expressed as

$$N(\lambda(\pi^T, y^{\mathbb{N}})) = \sum\nolimits_{i \in \mathbb{N}} \lambda_i(\pi^T, y^{\mathbb{N}})(Y \setminus \{y_0\}) = \sum\nolimits_{i \in \mathbb{N}} \sigma_i(\pi^T, y^{\mathbb{N}}) \tag{81}$$

Recalling that the utility $u(y_0)$ of non-existence is 0, person i's expected utility from the lottery $\lambda_i(\pi^T, y^{\mathbb{N}})$ in (79) is

$$\mathbb{E}_{\lambda_i(\pi^T, y^{\mathbb{N}})} u(y_i) = \pi_t \, u(y_i) = \sigma_i(\pi^T, y^{\mathbb{N}}) \, u(y_i) \tag{82}$$

Summing this over all $i \in \mathbb{N}$, it follows that the objective (68) takes the form

$$W(\lambda(\pi^T, y^{\mathbb{N}})) = \sum\nolimits_{i \in \mathbb{N}} \sigma_i(\pi^T, y^{\mathbb{N}}) \, u(y_i) \tag{83}$$

8.6 Extinction Discounting: Main Theorem

The following lemma relates the equations (81) and (83) to sums over time that incorporate appropriate extinction discounting.

Lemma 2. *Consider any extinction process* $\pi^T \in \Pi$ *and any personal consequence stream* $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ *with associated generation structure* $G^T(y^{\mathbb{N}})$*.*

1. *The expected population before extinction given by* (81) *satisfies*

$$N(\lambda(\pi^T, y^{\mathbb{N}})) = \sum\nolimits_{t \in T} \pi_t \, \# G_t(y^{\mathbb{N}}) \tag{84}$$

2. *Provided that* $N(\lambda(\pi^T, y^{\mathbb{N}}))$ *is finite, expected total utility is given by* (83)*, which is absolutely convergent, and satisfies*

$$W(\lambda(\pi^T, y^{\mathbb{N}})) = \sum\nolimits_{t \in T} \pi_t \sum\nolimits_{i \in G_t(y^{\mathbb{N}})} u(y_i) \tag{85}$$

Proof. By Definition 1 and Lemma 1 in Section 6.2, one has $\cup_{t \in T} G_t(y^{\mathbb{N}}) = \mathbb{N}$, where different sets $G_t(y^{\mathbb{N}})$ are pairwise disjoint. Now, all terms of the series (81) are non-negative. So, whether it converges or diverges to $+\infty$, one can rearrange these terms to obtain

$$N(\lambda(\pi^T, y^{\mathbb{N}})) = \sum\nolimits_{i \in \mathbb{N}} \sigma_i(\pi^T, y^{\mathbb{N}}) = \sum\nolimits_{t \in T} \sum\nolimits_{i \in G_t(y^{\mathbb{N}})} \sigma_i(\pi^T, y^{\mathbb{N}}) \tag{86}$$

But for all $i \in G_t(y^{\mathbb{N}})$, equations (79) and (80) imply $\sigma_i(\pi^T, y^{\mathbb{N}}) = \pi_t$, so

$$\sum\nolimits_{t \in T} \sum\nolimits_{i \in G_t(y^{\mathbb{N}})} \sigma_i(\pi^T, y^{\mathbb{N}}) = \sum\nolimits_{t \in T} \pi_t \, \# G_t(y^{\mathbb{N}}) \tag{87}$$

Together (86) and (87) confirm (84).

Next, provided that $N(\lambda(\pi^T, y^{\mathbb{N}}))$ is finite, Proposition 2 implies that the infinite series $\sum_{i \in \mathbb{N}} \mathbb{E}_{\lambda(\pi^T, y^{\mathbb{N}})} u(y_i)$ converges absolutely. By (82), so therefore does $\sum_{i \in \mathbb{N}} \sigma_i(\pi^T, y^{\mathbb{N}}) \, u(y_i)$. Rearranging the terms of this absolutely convergent series while using $\sigma_i(\pi^T, y^{\mathbb{N}}) = \pi_t$ for all $i \in G_t(y^{\mathbb{N}})$, it follows that

$$\sum\nolimits_{i \in \mathbb{N}} \sigma_i(\pi^T, y^{\mathbb{N}}) \, u(y_i) = \sum\nolimits_{t \in T} \pi_t \sum\nolimits_{i \in G_t(y^{\mathbb{N}})} u(y_i) \tag{88}$$

This proves that $W(\lambda(\pi^T, y^{\mathbb{N}}))$, which is given by (83), also satisfies (85). □

Lemma 2 motivates the following definition of extinction discounting, applied to both the population and the total utility of future generations.

Definition 8. *For any extinction process $\pi^T \in \Pi$ and any personal consequence sequence $y^{\mathbb{N}} \in Y^{\mathbb{N}}$ with associated generation structure $G^T(y^{\mathbb{N}})$:*

1. *the* extinction discounted total population *is given by* (84)*;*
2. *provided that the extinction discounted total population is finite,* extinction discounted total utility *is given by* (85)*.*

Finally, Definition 8 allows us to state our main result rather succinctly.

Theorem 1. *Suppose that the combination $(\pi^T, y^{\mathbb{N}})$ of an extinction process with a potential personal consequence stream together imply a finite extinction discounted total population given by* (84)*. Then the extinction discounted total utility given by* (85) *is well defined as an absolutely convergent infinite series.*

Proof. The result is an obvious implication of Lemma 2 and Definition 8. □

8.7 Reconciling Extinction Discounting with Sustainability

Superficially, the only difference between the welfare criterion (85) with extinction discounting and the criterion (47) with intergenerational discounting comes in the replacement of the discount factors β_t in the latter by the survival probabilities π_t in the former. This may make extinction discounting seem like a dictatorship of the present, just like discounted total utility. A crucial difference, however, is that dictatorship of the present does not seem so unreasonable when future generations may not exist.

Nevertheless, in order to retain sustainable preferences, one can adapt the criteria defined in (67) to the present setting with extinction discounting. The obvious result, for the domain of $(\pi^T, y^{\mathbb{N}})$ having the property that the extinction discounted expected total population given by (84) is finite, is the strictly convex combination

$$(1-\omega)\, W(\lambda(\pi^T, y^{\mathbb{N}})) + \omega \left[\alpha \liminf_{i\to\infty} u(y_i) + (1-\alpha) \limsup_{i\to\infty} u(y_i)\right] \qquad (89)$$

of the extinction discounting criterion (85) with a dictatorship of the future. Following Theorem 2 of Chichilnisky (1996) and the discussion in Section 7.7, an alternative is the sustainable von Neumann–Morgenstern utility function

$$(1-\omega)\, W(\lambda(\pi^T, y^{\mathbb{N}})) + \omega\, \Phi\left(\langle u(y_i)\rangle_{i\in\mathbb{N}}\right) \qquad (90)$$

Here $\ell^\infty \ni u^{\mathbb{N}} \mapsto \Phi(u^{\mathbb{N}}) \in \mathbb{R}$ is a suitable finitely additive measure that represents a continuous linear function on the space of bounded utility sequences in $\mathbb{R}^{\mathbb{N}}$.

It is interesting to consider what happens in case we depart from Definition 7 of Section 8.4, and allow ω to be the positive probability that extinction never occurs. Then both (89) and (90) represent total expected utility when π_t becomes the conditional probability of survival up to each time t, given that extinction occurs in finite time, and the term that multiplies ω measures social welfare in case extinction never occurs.

9 Conclusions

9.1 Revisiting Intergenerational Equity

We recall what Arrow (1999a, b) called the "strong argument" for discounting the welfare of future generations. This is that, given other standard assumptions, avoiding discounting would produce some form of logical contradiction. Our main conclusion, however, is that extinction discounting avoids any such contradiction under the following two assumptions: (i) the expected total population before extinction is finite; (ii) following Arrow (1951, 1965, 1971, 1972), the common fundamental utility function of each potential individual is bounded both above and below.

Our great friend Kenneth Arrow had espoused both strong and weak arguments in favour of discounting the welfare of future generations not only in writings such as Arrow (1999a, b, 2007), but also in later oral discussions with PJH and others. Yet his willingness to engage in these discussions suggests that, in the end, he may have been starting to experience some doubt about whether these two arguments, which he attributed to Koopmans, really had settled the discounting issue. Extinction discounting not only provides one possible escape from the logical contradiction that arises from trying to apply intergenerational equity to an infinite set of individuals. It may also allow an escape from the claim that too little discounting places too high a burden on current generations.

9.2 Beyond Welfarism

The welfarist approach that is most often used in public economics, including by KJA himself, is just one way of looking at the ethics of public and private decision-making. Indeed, several other approaches have commanded the attention of philosophers when they undertake ethical arguments. Among these others are: (i) contractarianism; (ii) Kantian ethics; (iii) Aristotelian ethics; (iv) arguments based on rights or liberty. This is not the place for a detailed discussion, some elements of which are provided in Stern (2014a;

2015, Chapter 6), along with references to further discussion of these different ethical perspectives.

Here we simply note that these four perspectives that go beyond welfarism would all seem to exclude discounting the welfare of individuals who belong to future generations. It follows that all urge strong action on climate change. Specifically, a contractarian would likely regard a reasonable or acceptable social contract as excluding decisions such as those resulting from an unacceptable dictatorship of the present that rides roughshod over future lives by refusing to manage climate change. A Kantian categorical imperative is to behave as you would have others behave; that would likely involve respecting others' livelihoods, even if they live decades later. Similarly, an Aristotelian notion of virtue would likely not permit causing serious damage to others' lives in pursuit of narrow self-interest. Finally, any approach based on rights would surely include respecting the rights of future generations.

We conclude that ethical arguments against discrimination by date of birth also apply in ethical frameworks that transcend utilitarian consequentialism. "Pure-time discounting" is not merely a narrow technical concern for nerdy economists. Avoiding this kind of discrimination is fundamental to most ethical doctrines.

9.3 A More Optimistic Scenario?

Much remains to be done in reforming the world's economic system in order to avoid the serious risk of catastrophic climate change due to excessive greenhouse gas emissions. Other serious risks include excessive acidification by dissolved CO_2, not only of oceans, but also of reserves of fresh water. Nevertheless, there are two reasons why the sacrifices required of either the current or future generations may turn out ultimately to be considerably less than had been feared back in the late 1990s. This was the period when negotiations were conducted which led to the Kyoto Protocol being approved by 160 nations in 1997, and becoming part of international law in 2005 after enough nations had ratified it.[27]

First, as is becoming widely recognized, the last ten or fifteen years have seen quite extraordinary technical progress, especially in using wind and solar power instead of steam to generate electricity. Along with technologies such as those that allow vehicles to use electric power to varying degrees, this is part of a general process whereby clean or zero-carbon technologies

[27]Chichilnisky and Sheeran (2009) offer one account of these negotiations, particularly their latter stages.

have become cheaper than high-carbon technologies in many sectors and geographical areas. In large measure this transformation has come about as a result of combining changes in social priorities with a process that KJA did so much to illuminate — namely the dynamics of learning by doing.[28]

Second, it may be appropriate to draw attention to the much less well known yet highly promising technologies for carbon dioxide removal by direct air capture that one of us (GC) has been helping to pioneer.[29] Moreover, there is some prospect in the next few years of a complementary technical change in the construction industry which, for instance, might allow the carbon-emitting activity of making cement to be replaced with the production of alternative stronger and lighter building materials based on some form of carbon fibre. If the carbon needed for this came from carbon dioxide directly captured from the air, it may yet prove possible to begin seriously reversing the build up of greenhouse gases in the atmosphere. Recently, a carbon X-prize was announced, intended to "challenge the world to reimagine what we can do with CO2 emissions by incentivizing and accelerating the development of technologies that convert CO2 into valuable products."[30]

Until relatively recently, it had seemed that the world would find it difficult to reduce the atmospheric concentration of greenhouse gases to safer levels without significant sacrifices of economic progress. The prospect of such sacrifices had become of great concern to KJA who, during his last two decades, addressed this kind of issue in several co-authored publications such as Arrow *et al.* (2003, 2004, 2012, 2013, 2014). Yet thanks in large part to recent technological developments, we are thinking more and more about policies that can manage change rather than discussing what "sacrifices" might be worthwhile. Indeed, there is the real prospect that directly capturing CO_2 from the air could soon replace existing sources of this widely used industrial gas and earn sufficient profit to make the sacrifice disappear entirely. In any case, this emphasis on how to manage change becomes even more urgent once we begin to consider other issues, such as those concerning human health and the robustness of ecosystems; here too KJA has had so much to say. His legacy is truly extraordinary.

[28]The work on the "learning curve" in Arrow (1962) was cited by the Stanford University School of Engineering when it recognized KJA as a "hero of engineering" — see `engineering.stanford.edu/about/heroes/kenneth-arrow`. The award occurred in the year 2014, whose other "heroes" included the two Stanford alumni who founded Google^{TM}, as well as the late Sally Ride, the first female U.S. astronaut.

[29]GC is the co-inventor and co-patentee of Carbon Negative Technology^{TM}, as described by Eisenberger, Cohen, Chichilnisky *et al.* (2009), Chichilnisky (2011), Chichilnisky and Eisenberger (2011), as well as Choi *et al.* (2011).

[30]See `carbon.xprize.org`.

Acknowledgements

For helpful discussions the authors would like to thank Kenneth Arrow in particular, with whom GC and PJH had the privilege of enjoying frequent academically productive and challenging lunch meetings during the last years of his long life. We also thank Peter Diamond, Juan Dolado, Humberto Llavador, James Mirrlees, and Eytan Sheshinski for helpful comments. In addition GC and NHS wish to credit PJH with doing more than his fair share of the work; PJH responds that therefore he is probably responsible for a disproportionate share of any errors.

References

Anscombe, G. Elizabeth M. (1958) "Modern Moral Philosophy" *Philosophy* 33: 1–19.

Arrow, Kenneth J. (1950) "A Difficulty in the Concept of Social Welfare" *Journal of Political Economy* 58 (4): 328–346.

Arrow, Kenneth J. (1951) "Alternative Approaches to the Theory of Choice in Risk-Taking Situations" *Econometrica* 19 (4): 404–437.

Arrow, Kenneth J. (1962) "The Economic Implications of Learning by Doing" *Review of Economic Studies* 29 (3): 155–173.

Arrow, Kenneth J. (1963) *Social Choice and Individual Values, 2nd edition* (Yale University Press).

Arrow, Kenneth J. (1965) *Aspects of the Theory of Risk-Bearing* (Amsterdam: North- Holland).

Arrow, Kenneth J. (1966) "Discounting and Public Investment Criteria" in A. V. Kneese and S. C. Smith (eds.) *Water Research* (Baltimore: Johns Hopkins University Press).

Arrow, Kenneth J. (1971) *Essays in the Theory of Risk-Bearing* (Chicago: Markham; Amsterdam: North-Holland).

Arrow, Kenneth J. (1972) "Exposition of the Theory of Choice under Uncertainty" in C.B. McGuire and R. Radner (eds.) *Decision and Organization* (Amsterdam: North-Holland), ch. 2, pp. 19–55.

Arrow, Kenneth J. (1973a) "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice" *Journal of Philosophy* 70, No. 9 (May 10): 245–263.

Arrow, Kenneth J. (1973b) "Rawls's Principle of Just Saving" *The Swedish Journal of Economics* 75 (4): 323–335.

Arrow, Kenneth J. (1977) "Extended Sympathy and the Possibility of Social Choice" *American Economic Review: Papers and Proceedings* 67 (1): 219–225.

Arrow, Kenneth J. (1982) "The Rate of Discount on Public Investments with Imperfect Capital Markets" in Lind, Robert C., Kenneth J. Arrow, Gordon R. Corey, Partha Dasgupta, Amartya K. Sen, Thomas Stauffer, Joseph E. Stiglitz, J.A. Stockfisch, and Robert Wilson (1982) *Discounting for Time and Risk in Energy Policy* (Washington, D.C.: Resources for the Future), ch. 4, pp. 115–136.

Arrow, Kenneth J. (1999a) "Inter-Generational Equity and the Rate of Discount in Long-Term Social Investment" In: Sertel, Murat R. (ed.) *Contemporary Economic Issues. International Economic Association Series.* (Palgrave Macmillan: London), pp. 89–102.

Arrow, Kenneth J. (1999b) "Discounting, Morality, and Gaming" in Portney, Paul R. and John P. Weyant (eds.) *Discounting and Intergenerational Effects* (Resources for the Future Press: Washington DC), pp. 13–21.

Arrow, Kenneth J. (2007) "Global Climate Change: A Challenge to Policy" *The Economists' Voice.* Published Online: 2007-06-30 DOI: `https://doi.org/10.2202/1553-3832.1270`.

Arrow, Kenneth J., William R. Cline, Karl-Göran Mäler, Mohan Munasinghe, and Joseph E. Stiglitz (1996) "Intertemporal Equity, Discounting, and Economic Efficiency" in J. P. Bruce, H. Lee, and E. F. Haites (eds.) *Climate change 1995: Economic and Social Dimensions of Climate Change, contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press).

Arrow, Kenneth J., Maureen L. Cropper, Christian Gollier, Ben Groom, Geoffrey M. Heal, Richard G. Newell, William D. Nordhaus, Robert S. Pindyck, William A. Pizer, Paul R. Portney, Thomas Sterner, Richard S. J. Tol and Martin L. Weitzman (2013) "Determining Benefits and Costs for Future Generations" *Science* 341: 349–350.

Arrow, Kenneth J., Maureen L. Cropper, Christian Gollier, Ben Groom, Geoffrey M. Heal, Richard G. Newell, William D. Nordhaus, Robert S. Pindyck, William A. Pizer, Paul R. Portney, Thomas Sterner, Richard S. J. Tol and Martin L. Weitzman (2014) "Should Governments Use a Declining Discount Rate in Project Analysis?" *Review of Environmental Economics and Policy* 8(2): 145–163.

Arrow, Kenneth J., Partha Dasgupta, Lawrence Goulder, Gretchen Daily, Paul Ehrlich, Geoffrey Heal, Simon Levin, Karl-Göran Mäler, Stephen Schneider, David Starrett, and Brian Walker (2004) "Are We Consuming Too Much?" *Journal of Economic Perspectives* 18 (3): 147–172.

Arrow, Kenneth J., Partha Dasgupta, Lawrence H. Goulder, Kevin J. Mumford, and Kirsten Oleson (2012) "Sustainability and the Measurement of Wealth" *Environment and Development Economics* 17 (3): 317–353.

Arrow, Kenneth J., Partha Dasgupta, and Karl-Göran Mäler (2003) "The Genuine Savings Criterion and the Value of Population" *Economic Theory* 21 (2): 217–225.

Arrow, Kenneth J., and Gérard Debreu (1954) "Existence of an Equilibrium for a Competitive Economy" *Econometrica* 22 (3): 265–290.

Arrow, Kenneth J., and Leonid Hurwicz (1972) "An Optimality Criterion for Decision-making Under Ignorance", in D. F. Carter and F. Ford (eds.) *Uncertainty and Expectations in Economics* (Oxford University Press).

Arrow, Kenneth J., and Mordecai Kurz (1970) *Public Investment, the Rate of Return and Optimal Fiscal Policy* (Baltimore: Johns Hopkins University Press).

Arrow, Kenneth J. and Robert C. Lind (1970) "Uncertainty and the Evaluation of Public Investment Decisions" *American Economic Review* 60 (3): 364–378.

Asheim, Geir B. (2007) "Economic Analysis of Sustainability" Chapter 1 of *Justifying, Characterizing and Indicating Sustainability* (Dordrecht: Springer).

Asheim, Geir B. (2010) "Intergenerational Equity" *Annual Review of Economics* 2: 197–222.

Asheim, Geir B., Tapan Mitra, and Bertil Tungodden (2012) “Sustainable Recursive Social Welfare Functions” *Economic Theory* 49 (2): 267–292.

Asheim, Geir B., and Stéphane Zuber (2013) “A Complete and Strongly Anonymous Leximin Relation on Infinite Streams” *Social Choice and Welfare* 41 (4): 819–834.

Atkinson, Anthony B. (2001) “The Strange Disappearance of Welfare Economics” *Kyklos* 54 (2/3): 193–206.

Basu, Kaushik, and Tapan Mitra (2003) “Aggregating Infinite Utility Streams With Intergenerational Equity: The Impossibility of Being Paretian” *Econometrica* 32: 1557–1563.

Basu, Kaushik, and Tapan Mitra (2007a) “Utilitarianism for Infinite Utility Streams: A New Welfare Criterion and Its Axiomatic Characterization” *Journal of Economic Theory* 133: 350–373.

Basu, Kaushik, and Tapan Mitra (2007b) “Possibility Theorems for Aggregating Infinite Utility Streams Equitably” In John Roemer and Kotaro Suzumura (eds.) *Intergenerational Equity and Sustainability* (Palgrave–Macmillan) ch. 5, pp. 69–84.

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

Bossert, Walter, Yves Sprumont and Kotaro Suzumura (2007) “Ordering Infinite Utility Streams” *Journal of Economic Theory* 135 (1): 579–589.

Bossert, Walter and Kotaro Suzumura (2011) “Multi-Profile Intergenerational Social Choice” *Social Choice and Welfare* 37 (3) : 493–509.

Brock, William A. (1970) “An Axiomatic Basis for the Ramsey–Weizsäcker Overtaking Criterion” *Econometrica* 38 (6): 927–929.

Broome, John (1993) “A Cause of Preference Is Not an Object of Preference” *Social Choice and Welfare* 10 (1): 57–68.

Broome, John (1994) “Discounting the Future” *Philosophy & Public Affairs* 23 (2): 128–156.

Brown, Campbell (2011) "Consequentialize This" *Ethics* 121 (4): 749–771.

Brown, Donald J., and Lewis, Lucinda M. (1981) "Myopic Economic Agents" *Econometrica* 49 (2): 359–368.

Brundtland, Gro Harlem (1987) *Report of the World Commission on Environment and Development: Our Common Future* (Oxford University Press, for the United Nations)

Burmeister, Edwin and Peter J. Hammond (1977) "Maximin Paths of Heterogeneous Capital Accumulation and the Instability of Paradoxical Steady States" *Econometrica* 45: 853–870.

Chichilnisky, Graciela (1977) "Economic Development and Efficiency Criteria in the Satisfaction of Basic Needs" *Applied Mathematical Modeling* 1 (6): 290–297.

Chichilnisky, Graciela (1996) "An Axiomatic Approach to Sustainable Development" *Social Choice and Welfare* 13 (2): 231–257.

Chichilnisky, Graciela (1997) "What is Sustainable Development?" *Land Economics* 73 (4): 467–491.

Chichilnisky, Graciela (2009) "Avoiding Extinction: Equal Treatment of the Present and the Future" *Economics: The Open-Access, Open-Assessment E-Journal* 3 (2009-32): 1–25. http://dx.doi.org/10.5018/economics-ejournal.ja.2009-32.

Chichilnisky, Graciela (2011) "Green Capitalism: Negative Carbon and the Green Power Fund" *International Journal of Green Economics*, Inderscience Enterprises Ltd, 5(4): 321–333.

Chichilnisky, Graciela, and Peter Eisenberger (2011) "Carbon Negative Power Plants" *CryoGas International*, May, p. 36.

Chichilnisky, Graciela, Geoffrey Heal and Andrea Beltratti (1995) "The Green Golden Rule" *Economics Letters* 49 (2): 175–179.

Chichilnisky, Graciela, and Kristen A. Sheeran (2009) *Saving Kyoto: An Insiders Guide to What It Is, How It Works and What It Means for the Future* (New Holland Publishers Ltd).

Choi, S., Drese, J.H, Eisenberger, P.M., and Jones, C.W. (2011) "Application of Amine-Tethered Solid Sorbents for Direct CO2 Capture from the Ambient Air" *Environmental Science & Technology* 45(6): 2420–7.

Creedy, John, and Ross Guest (2008) "Discounting and the Time Preference Rate" *Economic Record* 84 (264) : 109–127.

Dasgupta, Partha (1974) "On Some Problems Arising from Professor Rawls's Conception of Distributive Justice" *Theory and Decision* 4: 325–344.

Dasgupta, Partha (1994) "Savings and Fertility: Ethical Issues" *Philosophy & Public Affairs* 23 (2): 99–127.

Dasgupta, Partha (2007) "Commentary: The Stern Review's Economics of Climate Change" *National Institute Economic Review* 199: 4–7.

Dasgupta, Partha S. (2008) "Discounting Climate Change" *Journal of Risk and Uncertainty* 37 (2–3): 141–169.

Debreu, Gérard (1951) "The Coefficient of Resource Utilization" *Econometrica* 19 (3): 273–292.

Diamond, Peter A. (1965) "The Evaluation of Infinite Utility Streams" *Econometrica* 33: 170–177.

Dixit, Avinash, Peter J. Hammond and Michael Hoel (1980) "On Hartwick's Rule for Regular Maximin Paths of Capital Accumulation and Resource Depletion" *Review of Economic Studies* 47: 551–556.

Dubey, Ram Sewak, and Mitra, Tapan (2013) "On the Nature of Suppes–Sen Maximal Paths in an Aggregative Growth Model" *Social Choice and Welfare* 40: 173–205.

Edgeworth, Francis Ysidro (1925) *Papers Relating to Political Economy, Vol. III* (London: Macmillan).

Ehrlich, Paul R. (1968) *The Population Bomb* (Ballantine Books).

Eisenberger, Peter M., Cohen, R. W., Chichilnisky, G. *et al.* (2009) "Global Warming and Carbon-Negative Technology: Prospects for a Lower-Cost Route to a Lower-Risk Atmosphere" *Energy and Environment* 20: 973–984.

Figuières, Charles, and Mabel Tidball (2012) "Sustainable Exploitation of a Natural Resource: A Satisfying Use of Chichilnisky's Criterion" *Economic Theory* 49 (2): 243–265.

Fleurbaey, Marc, and Peter J. Hammond (2004) "Interpersonally Comparable Utility" in Salvador Barberà, Peter J. Hammond, and Christian Seidl (eds.) *Handbook of Utility Theory, Vol. 2: Extensions* (Kluwer Academic Publishers) ch. 21, pp. 1181–1285.

Fleurbaey, Marc, and Philippe Michel (2003) "Intertemporal Equity and the Extension of the Ramsey Criterion" *Journal of Mathematical Economics* 39: 777–802.

Gaertner, Wulf (2017) "Professor Kenneth Joseph Arrow (1921–2017)" *Social Choice and Welfare* 48 (4): 703–705.

Gale, David (1967) "On Optimal Development in a Multi-Sector Economy" *Review of Economic Studies* 34 (1): 1–18.

Hammond, Peter J. (1973) *Consistent Planning and Intertermporal Welfare Economics* Ph.D. dissertation, University of Cambridge.

Hammond, Peter J. (1974) "Overture Plans of Accumulation for an Infinite Horizon" University of Essex, Economics Discussion Paper No. 52 (1974).

Hammond, Peter J. (1975) "Agreeable Plans with Many Capital Goods" *Review of Economic Studies* 42 (1): 1–14.

Hammond, Peter J. (1986) "Consequentialist Social Norms for Public Decisions" in W.P. Heller, R.M. Starr and D.A. Starrett (eds.) *Social Choice and Public Decision Making: Essays in Honor of Kenneth J. Arrow, Vol. I* (Cambridge University Press), ch. 1, pp. 3–27.

Hammond, Peter J. (1988) "Consequentialist Demographic Norms and Parenting Rights" *Social Choice and Welfare* 5: 127–145.

Hammond, Peter J. (1991) "Interpersonal Comparisons of Utility: Why and how they are and should be made" in Jon Elster and John E. Roemer (eds.) *Interpersonal Comparisons of Well-Being* (Cambridge University Press), ch. 7, pp. 200–254.

Hammond, Peter J. (1992) "On the Impossibility of Perfect Capital Markets" in Partha Dasgupta, Douglas Gale, Oliver Hart, and Eric Maskin (eds.) *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn* (Cambridge, Mass.: M.I.T. Press), pp. 527–560.

Hammond, Peter J. (1993) "Is There Anything New in the Concept of Sustainable Development?" in L. Campiglio, L. Pineschi, D. Siniscalco, and T. Treves (eds.) *The Environment after Rio: International Law and Economics* (London: Graham & Trotman, 1994), ch. 13, pp. 185–194.

Hammond, Peter J. (1996a) "Consequentialism, Structural Rationality and Game Theory" in K.J. Arrow, E. Colombatto, M. Perlman, and C. Schmidt (eds.) *The Rational Foundations of Economic Behaviour* (IEA Conference Volume No. 114) (London: Macmillan) ch. 2, pp. 25–42.

Hammond, Peter J. (1996b) "Consequentialist Decision Theory and Utilitarian Ethics" in F. Farina, F. Hahn, and S. Vannucci (eds.) *Ethics, Rationality, and Economic Behaviour* (Oxford: Clarendon Press), pp. 92–118.

Hammond, Peter J. (1998) "Objective Expected Utility: A Consequentialist Perspective" in Salvador Barberà, Peter J. Hammond and Christian Seidl (ed.) *Handbook of Utility Theory, Vol. 1: Principles* (Kluwer Academic Publishers) ch. 5, pp. 145–211.

Hammond, Peter J., and John F. Kennan (1979) "Uniformly Optimal Infinite Horizon Plans" *International Economic Review* 20: 283–296.

Hammond, Peter J., and James A. Mirrlees (1973) "Agreeable Plans" in James A. Mirrlees and Nicholas H. Stern (eds.) *Models of Economic Growth* (London: Macmillan) ch. 13, pp. 283–299.

Hara, Chiaki, Tomoichi Shinotsuka, Kotaro Suzumura, and Yongsheng Xu (2008) "Continuity and Egalitarianism in the Evaluation of Infinite Utility Streams" *Social Choice and Welfare* 31 (2): 179–191.

Harrod, Roy. F. (1948) *Towards a Dynamic Economics* (London: Macmillan).

Harsanyi, John C. (1953) "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking" *Journal of Political Economy* 61: 434–435.

Harsanyi, John C. (1955) "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility" *Journal of Political Economy* 63 (4): 309–321.

Harsanyi, John C. (1975) "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory" *American Political Science Review* 69: 594–606.

Harsanyi, John C. (1977) "Morality and the Theory of Rational Behavior" *Social Research* 44: 623–656.

Harsanyi, John C. (1978) "Bayesian Decision Theory and Utilitarian Ethics" *American Economic Review (Papers and Proceedings)* 68: 223–228.

Harsanyi, John C. (1979) "Bayesian Decision Theory, Rule Utilitarianism, and Arrow's Impossibility Theorem" *Theory and Decision* 11: 289–317.

Hartwick, John M. (1977) "Intergenerational Equity and the Investment of Rents from Exhaustible Resources" *American Economic Review* 67: 972–74.

Heal, Geoffrey M. (1998) *Valuing the Future: Economic Theory and Sustainability* (Columbia Univ. Press)

Herrera, Amilcar Óscar, H. Scolnik, Graciela Chichilnisky *et al.* (1976) *Catastrophe or New Society?: A Latin American World Model (The Bariloche Model)* (Ottawa, Canada: International Development Research Center).

Hicks, John R. (1946) *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory* (Oxford: Clarendon Press).

Hume, David (1739) *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects* (London: John Noon).

Hutcheson, Francis (1725) *An Inquiry into the Original of our Ideas of Beauty and Virtue* (London).

Inagaki, Morido (1970) *Optimal Economic Growth. Shifting Finite Versus Infinite Time Horizon* (North-Holland).

Inagaki, Morido (1973) "Intertemporal National Optimality and Temporal Social Preferences" in James A. Mirrlees and Nicholas H. Stern (eds.) *Models of Economic Growth* (London: Macmillan) ch. 11, pp. 241–254.

Jonsson, Adam, and Mark Voorneveld (2018) "The Limit of Discounted Utilitarianism" *Theoretical Economics* 13: 19–37.

Kahneman, Daniel (2011) *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux).

Kolm, Serge-Christophe (1994) "The Meaning of 'Fundamental Preferences'" *Social Choice and Welfare* 11 (3): 193–198.

Koopmans, Tjalling C. (1960) "Stationary Ordinal Utility and Impatience" *Econometrica* 28 (2): 287–309.

Koopmans, Tjalling C. (1965) "On the Concept of Optimal Economic Growth" *Academiae Scientiarum Scripta Varia* 28 (1): 225–287; reprinted as Cowles Foundation Paper 238.

Koopmans, Tjalling C. (1967) "Objectives, Constraints, and Outcomes in Optimal Growth Models" *Econometrica* 35 (1): 1–15.

Koopmans, Tjalling C., Peter A. Diamond and Richard E. Williamson (1964) "Stationary Utility and Time Perspective" *Econometrica* 32 (1–2): 82–100.

Lauwers, Luc (1993) "Infinite Chichilnisky Rules" *Economics Letters* 42 (4): 349–352.

Lauwers, Luc (1997a) "Topological Aggregation, the Case of an Infinite Population" *Social Choice and Welfare* 14 (2): 319–332.

Lauwers, Luc (1997b) "Continuity and Equity With Infinite Horizons" *Social Choice and Welfare* 14 (2): 345–356.

Lauwers, Luc (1997c) "A Note on Weak ∞-Chichilnisky Rules" *Social Choice and Welfare* 14 (2): 357–359.

Lauwers, Luc (2010) "Ordering Infinite Utility Streams Comes at the Cost of a Non-Ramsey Set" *Journal of Mathematical Economics* 46: 32–37.

Lauwers, Luc (2017) "The Axiomatic Approach to the Ranking of Infinite Utility Streams" in Graciela Chichilnisky and Armon Rezai (eds.) *The Economics of the Global Environment: Catastrophic Risks in Theory and Policy* (Springer International 2016) pp. 231–255.

Little, Ian M.D., and James A. Mirrlees (1969) *Manual of Industrial Project Analysis in Developing Countries, 2* (Paris: OECD).

Llavador, Humberto, John E. Roemer, and Joaquim Silvestre (2010) “Intergenerational Justice when Future Worlds Are Uncertain” *Journal of Mathematical Economics* 46 (5): 728–761.

Llavador, Humberto, John E. Roemer, and Joaquim Silvestre (2011) “A Dynamic Analysis of Human Welfare in a Warming Planet” *Journal of Public Economics* 95 (11–12): 1607–1620.

Llavador, Humberto, John E. Roemer, and Joaquim Silvestre (2015) *Sustainability for a Warming Planet* (Cambridge, Mass: Harvard University Press).

Menger, Karl (1934) “Das Unsicherheitsmoment in der Wertlehre, Betrachtungen im Anschluss an das sogenannte Petersburger Spiel” *Zeitschrift für Nationalökonomie* 5: 459–485; translated under the title “The Role of Uncertainty in Economics” in M. Shubik (ed.) (1967) *Essays in Mathematical Economics in Honor of Oskar Morgenstern* (Princeton: Princeton University Press), ch. 16, pp. 211–231.

Mertens, Jean-François, and Anna Rubinchik (2012) “Intergenerational Equity and the Discount Rate for Cost–Benefit Analysis” *Macroeconomic Dynamics* 16(1), 61–93.

Mirrlees, James A. (1967) “Optimum Growth when Technology is Changing” *Review of Economic Studies* 34 (1): 95–124.

Mirrlees, James A., and Nicholas H. Stern (1972) “Fairly Good Plans” *Journal of Economic Theory* 4 (2): 268–288.

Mitra, Tapan, and Kaushik Basu (2007) “On the Existence of Paretian Social Welfare Relations for Infinite Utility Streams With Extended Anonymity” in John Roemer and Kotaro Suzumura (eds.) *Intergenerational Equity and Sustainability* (Basingstoke: Palgrave-Macmillan) ch. 6, pp. 85–99.

Mongin, Philippe (2001) “The Impartial Observer Theorem of Social Ethics” *Economics and Philosophy* 17: 147–179.

Mukerji, Nikil (2016) *The Case Against Consequentialism Reconsidered* (Springer).

Osumi, Keisuke (1986) *Economic Planning and Agreeability: An Investigation of Agreeable Plans in a General Class of Dynamic Economic Models* (Tankobon).

Parfit, Derek (1984) *Reasons and Persons* (Oxford University Press).

Pezzey, Jack (1997) "Sustainability Constraints Versus 'Optimality' Versus Intertemporal Concern, and Axioms Versus Data" *Land Economics* 73: 448–466.

Phelps, Edmund S., and Robert A. Pollak (1968) "On Second-Best National Saving and Game-Equilibrium Growth" *Review of Economic Studies* 35(2): 185–199.

Pigou, A.C. (1932) *The Economics of Welfare, Fourth Edition* (London: Macmillan).

Ponthière, Grégory (2003) "Should We Discount Future Generations' Welfare? A Survey on the 'Pure' Discount Rate Debate" University of Liège, Center of Research in Public Economics and Population Economics (CREPP), Working Paper 2003-02
`http://www2.ulg.ac.be/crepp/papers/crepp-wp200302.pdf`.

Portmore, Douglas W. (2007) "Consequentializing Moral Theories" *Pacific Philosophical Quarterly* 88: 39–73.

Portmore, Douglas W. (2009) "Consequentializing" *Philosophy Compass* 4: 329–347.

Quiggin, John (2012) "Equity Between Overlapping Generations" *Journal of Public Economic Theory* 14 (2): 273–283.

Ramsey, Frank P. (1928) "A Mathematical Theory of Saving" *Economic Journal* 38 (4): 543–549.

Ramsey, Frank P. (1931) "Epilogue" in R.B. Braithwaite (ed.) *F.P. Ramsey, The Foundations of Mathematics and Other Logical Essays* (London: Routledge and Kegan Paul), pp. 287–292.

Rawls, John (1971) *A Theory of Justice* (Belknap Press).

Roberts, Kevin W.S. (1980) "Interpersonal Comparability and Social Choice Theory" *Review of Economic Studies* 47: 421–439.

Roemer, John E. (2011) "The Ethics of Intertemporal Distribution in a Warming Planet" *Environmental and Resource Economics* 48 (3): 363–390.

Savage, Leonard J. (1954) *The Foundations of Statistics* (New York: John Wiley).

Sen, Amartya K. (1970, 2017) *Collective Choice and Social Welfare: Expanded Edition* (Penguin Books).

Sidgwick, Henry (1907) *The Methods of Ethics* (Macmillan: London).

Solow, Robert M. (1956) "A Contribution to the Theory of Economic Growth" *Quarterly Journal of Economics* 70 (1): 65–94.

Solow, Robert M. (1974) "Intergenerational Equity and Exhaustible Resources" *Review of Economic Studies* 41 (Symposium): 29–45.

Solow, Robert M. (1991) "Sustainability: An Economist's Perspective". The Eighteenth J. Seward Johnson Lecture, Marine Policy Center, Woods Hole Oceanographic Institution.

Solow, Robert (2012) "A Few Comments on 'Sustainability and the Measurement of Wealth'" *Environment and Development Economics* 17 (3): 354–355.

Stern, Nicholas H. (1977) "Welfare Weights and the Elasticity of the Marginal Valuation of Income" in M. Artis and R. Nobay (eds.) *Proceedings of the AUTE Edinburgh Meeting of April 1976* (Basil Blackwell) ch. 8.

Stern, Nicholas H. (2006) *The Economics of Climate Change: The Stern Review* (Cambridge University Press).

Stern, Nicholas H. (2008) "The Economics of Climate Change" *American Economic Review* 98 (2): 1–37.

Stern, Nicholas H. (2013) "The Structure of Economic Modeling of the Potential Impacts of Climate Change: Grafting Gross Underestimation of Risk onto Already Narrow Science Models" *Journal of Economic Literature* 51(3): 838–859.

Stern, Nicholas H. (2014a) "Ethics, Equity and the Economics of Climate Change. Paper 1: Science and Philosophy" *Economics and Philosophy* 30 (3): 397–444.

Stern, Nicholas H. (2014b) "Ethics, Equity and the Economics of Climate Change. Paper 2: Economics and Politics" *Economics and Philosophy* 30 (3): 445–501.

Stern, Nicholas H. (2015) *Why Are We Waiting? The Logic, Urgency, and Promise of Tackling Climate Change* (MIT Press).

Suppes, Patrick C. (1966) "Some Formal Models of Grading Principles" *Synthese* 6: 284–306.

Svensson, Lars-Gunnar (1980) "Equity Among Generations" *Econometrica* 48 (5): 1251–1256.

Swan, Trevor W. (1956) "Economic Growth and Capital Accumulation" *Economic Record* 32: 334–361.

Tinbergen, Jan (1957) "Welfare Economics and Income Distribution" *American Economic Review: Papers and Proceedings* 47 (2): 490–503.

Vickrey, William S. (1945) "Measuring Marginal Utility by Reactions to Risk" *Econometrica* 13 (4): 319–333.

Vickrey, William S. (1960) "Utility, Strategy and Social Decision Rules" *Quarterly Journal of Economics* 74 (4): 507–535.

von Weizsäcker, Carl Christian (1965) "Existence of Optimal Programs of Accumulation for an Infinite Time Horizon" *Review of Economic Studies* 32 (1): 85–104.

Yaari, Menahem E. (1981) "Rawls, Edgeworth, Shapley, Nash: Theories of Distributive Justice Re-examined" *Journal of Economic Theory* 24 (1): 1–39.

Zame, William (2007) "Can Intergenerational Equity Be Operationalized?" *Theoretical Economics* 2: 187–202.

Zuber, Stéphane, and Geir B. Asheim (2012) "Justifying Social Discounting: The Rank-Discounted Utilitarian Approach" *Journal of Economic Theory* 147 (4): 1572–1601.