Sustainable preferences via nondiscounted, hyperreal intergenerational welfare functions

Pivato, Marcus

Department of Mathematics, Trent University

5 March 2008

Sustainable preferences via nondiscounted, hyperreal intergenerational welfare functions

Marcus Pivato
Department of Mathematics, Trent University
Peterborough, Ontario, Canada
marcuspivato@trentu.ca

March 5, 2008

Abstract

We define an intergenerational social welfare function ${}^*\Sigma$ from $\mathbb{R}^{\mathbb{N}}$ (the set of all infinite-horizon utility streams) into ${}^*\mathbb{R}$ (the ordered field of hyperreal numbers). The function ${}^*\Sigma$ is continuous, linear, and increasing, and is well-defined even on unbounded (e.g. exponentially increasing) utility streams. This yields a complete social welfare ordering on $\mathbb{R}^{\mathbb{N}}$ which is Pareto and treats all generations equally (i.e. does not discount future utility). In particular, it is what Chichilnisky (1996) calls a 'sustainable' preference ordering: it is neither a 'dictatorship of the present' nor a 'dictatorship of the future'.

We then show how an agent with no 'pure' time preferences may still 'informationally discount' the future, due to uncertainty. Last, we model intergenerational choice for an exponentially growing economy and population. In one parameter regime, our model shows 'instrumental discounting' due to declining marginal utility of wealth. In another regime, we see a disturbing 'Paradox of Eternal Deferral'.

If the consequences of economic decisions unfold over time, then these decisions require tradeoffs between payoffs at one time and payoffs at later times. A rational economic agent must have some systematic way to evaluate these tradeoffs; this is the problem of *intertemporal choice*. Likewise, society as a whole must make long-term plans (e.g. investments in infrastructure or environmental protection) which affect the welfare of future generations; this is the problem of *intergenerational choice*.

In most economic models, each agent makes intertemporal choices by maximizing a *discounted sum* of future expected utilities. Formally, the agent fixes some *discounting sequence*[1] $\mathbf{d} = (d_t)_{t=0}^{\infty} \in \ell^1$, where $\ell_1 := \left\{\mathbf{d} \in \mathbb{R}^{\mathbb{N}}_{\neq} \,;\, \sum_{t=0}^{\infty} d_t < \infty\right\}$. The 'present value'

[1] Normally, $d_t := \lambda^t$, where $\lambda \in (0,1)$ is some *discount factor*. Indeed, such exponential discounting is required for time-consistency (Caplin and Leahy, 2004).

of any expected utility stream $\mathbf{u} = (u_t)_{t=0}^{\infty} \in \mathbb{R}^{\mathbb{N}}$ is then defined by $\langle \mathbf{d}, \mathbf{u} \rangle := \sum_{t=0}^{\infty} d_t u_t$. The agent chooses the strategy with the maximal present value.

While this discounted sum might be appropriate for the intertemporal choice by an impatient individual, it is arguably inappropriate as an intergenerational welfare function (IGWF) for a society, because it systematically discriminates against future generations. Indeed, Cowen and Parfit (1992), Chichilnisky (1996) and others have argued that the inappropriate application of exponential discounting to intergenerational choice would lead to environmentally unsustainable policies: if we applied a 3% discount rate (which is typical for an individual), then even catastrophic long-term environmental consequences (e.g. due to global warming) would be 'discounted' into insignificance, and hence, would have no influence on present-day economic planning. Chichilnisky (1996) calls such a myopic IGWF a 'dictatorship of the present' . However, people are in fact quite concerned about global warming; this indicates that this is *not* how they implicitly think about intergenerational choice.

Chichilnisky argues that we need a different IGWF, which properly accounts for the far-future consequences of our actions. However, it is also inappropriate for the IGWF to focus entirely on long-term consequences (e.g. to maximize $\liminf_{t\to\infty} u_t$), while ignoring short-term consequences; Chichilnisky (1996) calls such an IGWF a 'dictatorship of the future'. Instead, we need a balanced approach (she calls this a 'sustainable' IGWF). Let $\mathbb{U} := [-1, 1]^{\infty}$, so that $\mathbb{U}^{\mathbb{N}}$ is the space of all utility streams uniformly bounded by 1; Chichilnisky shows that, if $\chi : \mathbb{U}^{\mathbb{N}} \longrightarrow \mathbb{R}$ is a continuous, linear IGWF, then χ is sustainable if and only if

$$\forall\ \mathbf{u} \in \mathbb{U}^{\mathbb{N}}, \qquad \chi(\mathbf{u}) \quad = \quad \langle \mathbf{d}, \mathbf{u} \rangle + \int_{\mathbb{N}} \mathbf{u}\, d\phi, \tag{1}$$

where $\mathbf{d} = (d_t)_{t=0}^{\infty} \in \ell^1$ is some summable discounting sequence, and $\phi : \mathcal{P}(\mathbb{N}) \longrightarrow \mathbb{R}$ is a purely finitely additive (PFA) measure on $\mathbb{N}$. A PFA measure assign zero mass to all finite subsets of $\mathbb{N}$; thus, $\int_{\mathbb{N}} \mathbf{u}\, d\phi$ is only sensitive to the asymptotic properties of the sequence $(u_t)_{t=0}^{\infty}$ as $t\to\infty$. For example, Chichilnisky suggests that we use a PFA measure obtained by defining $\int_{\mathbb{N}} \mathbf{u}\, d\phi := \lim_{t\to\infty} u_t$ whenever this limit exists, and then extending $\int\ d\phi$ to all of $\mathbb{U}^{\mathbb{N}}$ via the Hahn-Banach Theorem.

However, Chichilnisky's function χ still has some shortcomings. First of all, although it avoids dictatorship of the present and of the future, χ still does not treat all generations equally: near-future generations are favoured by the discounted sum $\langle \mathbf{d}, \bullet \rangle$, whereas extreme far-future generations are favoured by the PFA measure ϕ. Intermediate generations are favoured by neither. In short: χ is not invariant under permutation of generations. Another problem is that χ is only well-defined for uniformly bounded utility streams. However, intergenerational choice must allow for the possible long-term growth of either the population, or per capita income, or both; hence we must allow for the possibility that the (population-weighted) aggregate expected utility stream $(u_t)_{t=0}^{\infty}$ grows without bound (perhaps even exponentially). Thus, to be useful, an IGWF must be able to compare two unbounded utility streams.

In §1 we will develop such an IGWF, and in §4 we will apply it to an economy with exponential growth. Before that, in §2 we briefly discuss how discounting can arise *without* pure time preferences. We elaborate on this idea in §3, where we consider ‘informational discounting’ due to uncertainty.

1 A hyperreal, nondiscounted sum of future utilities

A *free ultrafilter* on $\mathbb{N}$ is a collection $\mathcal{F}$ of subsets of $\mathbb{N}$ such that:

(a) If $\mathbf{A}, \mathbf{B} \in \mathcal{F}$, then $\mathbf{A} \cap \mathbf{B} \in \mathcal{F}$.

(b) If $\mathbf{A} \in \mathcal{F}$ and $\mathbf{A} \subseteq \mathbf{B}$, then $\mathbf{B} \in \mathcal{F}$.

(c) If $\mathbf{A}$ is finite, then $\mathbf{A} \notin \mathcal{F}$. (In particular, $\emptyset \notin \mathcal{F}$).

(d) For any $\mathbf{A} \subseteq \mathbb{N}$, either $\mathbf{A}$ or $\mathbf{A}^{\complement}$ is in $\mathcal{F}$ [but not both, because of (a) and (c)].

It helps to visualize this topologically. The *Stone-Čech compactification* of $\mathbb{N}$ is a compact Hausdorff space $\beta\mathbb{N}$ along with a dense embedding $\mathbb{N} \hookrightarrow \beta\mathbb{N}$; in a certain sense $\beta\mathbb{N}$ is the ‘largest’ compact Hausdorff space into which $\mathbb{N}$ can be densely embedded. Thus, elements of $(\beta\mathbb{N}) \setminus \mathbb{N}$ are cluster points of $\mathbb{N}$ which lie just outside of $\mathbb{N}$ itself; they are like ‘cluster points at infinity’. If $\mathcal{F}$ is any free ultrafilter, then there is a unique $\widetilde{n} \in (\beta\mathbb{N}) \setminus \mathbb{N}$ such that $\mathcal{F}$ is the set of all neighbourhoods of $\widetilde{n}$. If $\mathbf{F} \in \mathcal{F}$, and some property holds for all $n \in \mathbf{F}$, then this property holds ‘in the limit’ as $n \to \widetilde{n}$.

Let $\mathcal{F}$ be a free ultrafilter on $\mathbb{N}$, and define the relation $\underset{\mathcal{F}}{\approx}$ on $\mathbb{R}^{\mathbb{N}}$ by $\mathbf{x} \underset{\mathcal{F}}{\approx} \mathbf{y}$ iff $\{n \in \mathbb{N}\,;\; x_n = y_n\} \in \mathcal{F}$. Then ${}^*\mathbb{R} := \mathbb{R}^{\mathbb{N}} / \underset{\mathcal{F}}{\approx}$ is the set of *hyperreal* (or *nonstandard real*) *numbers*.[2] For any $\mathbf{x} \in \mathbb{R}^{\mathbb{N}}$, let $[\mathbf{x}] \in {}^*\mathbb{R}$ be its $\underset{\mathcal{F}}{\approx}$-equivalence class. There is a natural embedding $\mathbb{R} \ni r \mapsto {}^*r \in {}^*\mathbb{R}$, where ${}^*r := [(r, r, r, \ldots)]$. For any $[\mathbf{x}], [\mathbf{y}] \in {}^*\mathbb{R}$, we write $[\mathbf{x}] \underset{*}{\leq} [\mathbf{y}]$ if $\{n \in \mathbb{N}\,;\; x_n \leq y_n\} \in \mathcal{F}$. Then $\underset{*}{\leq}$ is well-defined, and defines a total ordering on ${}^*\mathbb{R}$. We define $[\mathbf{x}] \underset{*}{+} [\mathbf{y}] := [(x_0 + y_0,\, x_1 + y_1,\, x_2 + y_2,\, \ldots)]$, and $\underset{*}{-} [\mathbf{x}] := [(-x_0, -x_1, -x_2, \ldots)]$. We can define multiplication and division similarly; then ${}^*\mathbb{R}$ is a totally ordered field, and the embedding $\mathbb{R} \longrightarrow {}^*\mathbb{R}$ is a monomorphism.

For any $\mathbf{u} := (u_t)_{t=0}^{\infty} \in \mathbb{R}^{\mathbb{N}}$, we define ${}^*\sum \mathbf{u} := [\mathbf{x}]$, where $\mathbf{x} \in \mathbb{R}^{\mathbb{N}}$ is defined by $x_n := \displaystyle\sum_{t=0}^{n} u_t$ for all $n \in \mathbb{N}$. This defines a linear function ${}^*\sum : \mathbb{R}^{\mathbb{N}} \longrightarrow {}^*\mathbb{R}$. For example, if $(u_t)_{t=0}^{\infty}$ is an expected utility stream, then ${}^*\sum \mathbf{u}$ is the *nondiscounted* total lifetime utility. Such a nondiscounted sum generally does not converge to a real number, but it *does* converge to a well-defined hyperreal number. Thus, if $\mathbb{N}$ indexes an infinite sequence of generations, so that u_t represents the expected aggregate utility of

[2] Up to isomorphism, this definition does not depend on the choice of ultrafilter $\mathcal{F}$. See Rashid (1987), Anderson (1991) or Arkeryd et al. (1997) for introductions to hyperreal numbers and nonstandard analysis.

generation t, then the function ${}^*\sum : \mathbb{R}^{\mathbb{N}} \longrightarrow {}^*\mathbb{R}$ is an intergenerational welfare function (IGWF). This leads to a complete intergenerational welfare ordering $\preceq$ on $\mathbb{R}^{\mathbb{N}}$. For any utility streams $\mathbf{u} = (u_t)_{t=0}^{\infty}$ and $\mathbf{v} = (v_t)_{t=0}^{\infty} \in \mathbb{R}^{\mathbb{N}}$,

$$\Big(\mathbf{u} \preceq \mathbf{v}\Big) \iff \Big({}^*\textstyle\sum \mathbf{u} \underset{*}{\leq} {}^*\sum \mathbf{v}\Big) \iff \left(\left\{n \in \mathbb{N}\ ;\ \displaystyle\sum_{t=0}^{n} u_t \leq \sum_{t=0}^{n} v_t\right\} \in \mathcal{F}\right).$$

This is similar to the 'overtaking' criterion of von Weizacker (1967), but restricted to the ultrafilter $\mathcal{F}$, so that it defines a complete ordering on $\mathbb{R}^{\mathbb{N}}$. As an intergenerational welfare function, ${}^*\sum$ has several nice properties.

Pareto property. If $u_t \geq v_t$ for all $t \in \mathbb{N}$, and $u_t > v_t$ for at least one $t \in \mathbb{N}$, then clearly ${}^*\sum \mathbf{u} > {}^*\sum \mathbf{v}$. In particular, ${}^*\sum$ is strictly increasing in each coordinate.

Intergenerational egalitarianism. ${}^*\sum$ treats all generations equally. To be precise, let $\sigma : \mathbb{N} \longrightarrow \mathbb{N}$ be a bijection, and let $\mathbf{S} := \{n \in \mathbb{N}\ ;\ \sigma([1...n]) = [1...n]\}$. We say σ is an ***$\mathcal{F}$-semifinite permutation*** if $\mathbf{S} \in \mathcal{F}$. For example:

- Suppose $\sigma(t) = t$ for all but finitely many $t \in \mathbb{N}$; then σ is $\mathcal{F}$-semifinite.
- Let $\mathbf{E} := \{0, 2, 4, \ldots\}$ and $\mathbf{O} := \{1, 3, 5, \ldots\}$ be the sets of even and odd numbers; then either $\mathbf{E} \in \mathcal{F}$ or $\mathbf{O} \in \mathcal{F}$ (but not both). If $\mathbf{O} \in \mathcal{F}$, then define $\sigma(n) := n-1$ if $n \in \mathbf{O}$ and $\sigma(n) := n + 1$ if $n \in \mathbf{E}$; then σ is $\mathcal{F}$-semifinite (with $\mathbf{S} = \mathbf{O}$). If $\mathbf{E} \in \mathcal{F}$, then let $\sigma(0) := 0$, and for all $n \geq 1$, define $\sigma(n) := n - 1$ if $n \in \mathbf{E}$ and $\sigma(n) := n + 1$ if $n \in \mathbf{O}$; then σ is again $\mathcal{F}$-semifinite (with $\mathbf{S} = \mathbf{E}$). In either case, we have $\sigma(\mathbf{O}) \subseteq \mathbf{E}$ and $\sigma(\mathbf{E} \setminus \{0\}) \subseteq \mathbf{O}$.

Let $\mathbf{u} := (u_t)_{t=0}^{\infty}$, and define $v_t := u_{\sigma(t)}$ for all $t \in \mathbb{N}$. Then ${}^*\sum \mathbf{v} = {}^*\sum \mathbf{u}$ (because for any $n \in \mathbf{S}$, we have $\sum_{t=0}^{n} v_t = \sum_{t=0}^{n} u_t$, and $\mathbf{S} \in \mathcal{F}$). In other words, ${}^*\sum$ is invariant under any $\mathcal{F}$-semifinite permutation of generations. (Note that neither the discounted sum $\langle \mathbf{d}, \bullet \rangle$ nor the function χ in eqn.(1) have this property).

Handles exponential economic growth. ${}^*\sum \mathbf{u}$ is well-defined even if the utility stream $(u_t)_{t=0}^{\infty}$ grows without bound. In contrast, the function χ in eqn.(1) requires $\mathbf{u}$ to be uniformly bounded. The discounted sum $\langle \mathbf{d}, \bullet \rangle$ also has problems: if the exponential growth rate of $\mathbf{u}$ exceeds the discount rate of $\mathbf{d}$, then $\langle \mathbf{d}, \mathbf{u} \rangle = \infty$.

Infinitesimal impatience. The discounted sum $\langle \mathbf{d}, \bullet \rangle$ is 'impatient' in the sense that it prefers immediate payoffs to future ones. Since ${}^*\sum$ is a *non*discounted sum of future utilities, it exhibits no 'real' impatience; however it does exhibit some 'infinitesimal' impatience, as follows. Let $\mathbf{u} = (u_t)_{t=0}^{\infty}$ and $\mathbf{v} = (v_t)_{t=0}^{\infty}$ and $R \in \mathbb{R}$ be such that

$$\sum_{t=0}^{\infty} u_t \;=\; R \;=\; \sum_{t=0}^{\infty} v_t, \quad \text{(where} \quad \sum_{t=0}^{\infty} u_t \;:=\; \lim_{T\to\infty} \sum_{t=0}^{T} u_t, \quad \text{etc., as usual).} \qquad (2)$$

Suppose, however, that $\sum_{t=0}^{T} u_t < \sum_{t=0}^{T} v_t$ for all $T \in \mathbb{N}$. (For example, suppose $(v_t)_{t=0}^{\infty} \in \mathbb{R}_+^{\mathbb{N}}$ is a strictly positive sequence, and $u_0 := 0$ and $u_t := v_{t-1}$ for all $t > 0$; then $\sum_{t=0}^{T} u_t = \sum_{t=0}^{T-1} v_t < \sum_{t=0}^{T} v_t$.)

Equation (2) suggests that ${}^*\sum \mathbf{u} = {}^*R = {}^*\sum \mathbf{v}$, but this is not the case. If $\sum_{t=0}^{T} u_t < \sum_{t=0}^{T} v_t$ for all T, then it is easy to see that ${}^*\sum \mathbf{u} \underset{*}{<} {}^*\sum \mathbf{v}$. The difference $\epsilon := {}^*\sum \mathbf{v} - {}^*\sum \mathbf{u}$ is an *infinitesimal*; that is, ϵ is a hyperreal number such that ${}^*0 \underset{*}{<} \epsilon \underset{*}{<} {}^*r$ for any real number $r > 0$. Thus, ${}^*\sum$ *does* prefer the expected utility stream $\mathbf{v}$ over $\mathbf{u}$ by an infinitesimal amount, reflecting the fact that, although both streams deliver the same total expected utility, $\mathbf{v}$ delivers it slightly sooner.

Continuity. The function ${}^*\sum : \mathbb{R}^{\mathbb{N}} \longrightarrow {}^*\mathbb{R}$ is not continuous in the (Tychonoff) product topology on $\mathbb{R}^{\mathbb{N}}$ —indeed, there is *no* function from $\mathbb{R}^{\mathbb{N}}$ to ${}^*\mathbb{R}$ which is Pareto, invariant under finite coordinate permutations, and continuous in the product topology (Efimov and Koshevoy, 1994; Lauwers, 1997a). Neither is ${}^*\sum$ continuous in the Mackey topology. In fact, as Bewley (1970) observed, if a preference ordering on $\mathbb{R}^{\mathbb{N}}$ is continuous in the product or Mackey topologies, then it must be 'myopic' in a certain precise sense. [The set of all such 'myopia-inducing' topologies on $\mathbb{R}^{\mathbb{N}}$ was characterized by Brown and Lewis (1981).]

Also, unlike Chichilnisky's IGWF χ in eqn.(1), ${}^*\sum$ is not continuous in the ℓ^{∞} norm topology on $[-1,1]^{\mathbb{N}}$. However, ${}^*\sum$ does satisfy a strictly weaker form of continuity. Let $d_1 : \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}} \longrightarrow \mathbb{R} \cup \{\infty\}$ be the ℓ_1 pseudometric $d_1(\mathbf{u},\mathbf{v}) := \sum_{t=0}^{\infty} |u_t - v_t|$ ('pseudo' because $d(\mathbf{u},\mathbf{v})$ could be infinite). Let $d_e : {}^*\mathbb{R} \times {}^*\mathbb{R} \longrightarrow \mathbb{R}$ be the Euclidean metric[3]: $d_e(x,y) := \inf\{r \in \mathbb{R} \; ; \; -r < x - y < r\}$. Then ${}^*\sum$ is (d_1, d_e)-continuous.

No dictatorship of the present or the future. Let $\mathbb{U} \subseteq \mathbb{R}$ be a set of admissible utility levels (e.g. in Chichilnisky (1996), $\mathbb{U} := [-1,1]$; in our model, $\mathbb{U} := \mathbb{R}$). For any $\mathbf{u} = (u_t)_{t=0}^{\infty} \in \mathbb{U}^{\mathbb{N}}$ and $T \in \mathbb{N}$, let $\mathbf{u}^T := (u_t)_{t=0}^{T}$ and $\mathbf{u}_T := (u_t)_{t=T+1}^{\infty}$. If $\mathbf{v} = (v_t)_{t=0}^{\infty} \in \mathbb{U}^{\mathbb{N}}$, let $(\mathbf{u}^T, \mathbf{v}_T) := (w_t)_{t=0}^{\infty}$, where $w_t := u_t, \; \forall \, t \leq T$ and $w_t := v_t, \; \forall \, t > T$.

If $\Phi : \mathbb{U}^{\mathbb{N}} \longrightarrow \mathbb{R}$ is an IGWF, and $\mathbf{u}, \mathbf{v} \in \mathbb{U}^{\mathbb{N}}$, then Φ *myopically prefers* $\mathbf{u}$ to $\mathbf{v}$ if $\Phi(\mathbf{u}) > \Phi(\mathbf{v})$ and there is some $T \in \mathbb{N}$ such that, for *any* $\widetilde{\mathbf{u}}, \widetilde{\mathbf{v}} \in \mathbb{U}^{\mathbb{N}}$, we have $\Phi(\mathbf{u}^T, \widetilde{\mathbf{u}}_T) > \Phi(\mathbf{v}^T, \widetilde{\mathbf{v}}_T)$. In other words, the fact that $\mathbf{u}$ is socially preferred to $\mathbf{v}$ is entirely determined by the short-term structure of $\mathbf{u}$ and $\mathbf{v}$ —their long-term properties are irrelevant. Chichilnisky (1996) calls Φ a 'dictatorship of the present' if, for all $\mathbf{u}, \mathbf{v} \in \mathbb{U}^{\mathbb{N}}$, if $\Phi(\mathbf{u}) > \Phi(\mathbf{v})$ then Φ myopically prefers $\mathbf{u}$ to $\mathbf{v}$. For example, if $\mathbf{d} \in \ell^1$ is any summable discount sequence, and $\mathbb{U}$ is a bounded set, then the IGWF $\langle \mathbf{d}, \bullet \rangle$ is a dictatorship of the present (Chichilnisky, 1996, Thm.1). (Note that this is false if $\mathbb{U}$ is unbounded). We will actually weaken Chichilnisky's definition somewhat: If $\Phi : \mathbb{U}^{\mathbb{N}} \longrightarrow {}^*\mathbb{R}$ is an IGWF, we will say that Φ is a *weak dictatorship of the present* if there *exist* some $\mathbf{u}, \mathbf{v} \in \mathbb{U}^{\mathbb{N}}$ such that $\Phi(\mathbf{u}) \underset{*}{>} \Phi(\mathbf{v})$ and Φ myopically prefers $\mathbf{u}$ to $\mathbf{v}$.

[3]Note for each $x \in \mathbb{R}$ there exist points $y \in {}^*\mathbb{R}$ such that $x \neq y$ but $d_e(x,y) = 0$ —i.e. y is 'infinitesimally close' to x. This is one of the peculiarities of nonstandard topology.

If $\Phi : \mathbb{U}^{\mathbb{N}} \longrightarrow \mathbb{R}$ is an IGWF, and $\mathbf{u}, \mathbf{v} \in \mathbb{U}^{\mathbb{N}}$, then Φ *eternally prefers* $\mathbf{u}$ to $\mathbf{v}$ if $\Phi(\mathbf{u}) > \Phi(\mathbf{v})$ and there is some $T \in \mathbb{N}$ such that, for *any* $\widetilde{\mathbf{u}}, \widetilde{\mathbf{v}} \in \mathbb{U}^{\mathbb{N}}$, we have $\Phi(\widetilde{\mathbf{u}}^T, \mathbf{u}_T) > \Phi(\widetilde{\mathbf{v}}^T, \mathbf{v}_T)$. In other words, the fact that $\mathbf{u}$ is socially preferred to $\mathbf{v}$ is entirely determined by the long-term structure of $\mathbf{u}$ and $\mathbf{v}$ —their short-term properties are irrelevant. Chichilnisky (1996) calls Φ a 'dictatorship of the future' if for all $\mathbf{u}, \mathbf{v} \in \mathbb{U}^{\mathbb{N}}$, if $\Phi(\mathbf{u}) > \Phi(\mathbf{v})$ then Φ eternally prefers $\mathbf{u}$ to $\mathbf{v}$. For example, $\Phi(\mathbf{u}) := \liminf_{t\to\infty} u_t$ is a dictatorship of the future (Chichilnisky, 1996, Thm.1).

We must modify Chichilnisky's definition slightly to account for the possibility of hyperreal social welfare. If $r \in {}^*\mathbb{R}$, then r is *hyperfinite* if $r \underset{*}{>} {}^*n$ for all $n \in \mathbb{N}$, or if $-r \underset{*}{>} {}^*n$ for all $n \in \mathbb{N}$. Otherwise, r is *finite*. Thus, all real numbers are finite, and a real number plus an infinitesimal hyperreal is still finite.

If $\Phi : \mathbb{U}^{\mathbb{N}} \longrightarrow {}^*\mathbb{R}$ is a (hyperreal) IGWF, then we say that Φ is *weak dictatorship of the future* if there exist $\mathbf{u}, \mathbf{v} \in \mathbb{U}^{\mathbb{N}}$ with $\Phi(\mathbf{u}) \underset{*}{>} \Phi(\mathbf{v})$, such that $\Phi(\mathbf{u}) \underset{*}{-} \Phi(\mathbf{v})$ finite, but Φ eternally prefers $\mathbf{u}$ to $\mathbf{v}$. Note that, if $\Phi(\mathbf{u}) \underset{*}{-} \Phi(\mathbf{v})$ is hyperfinite, then Φ may eternally prefer $\mathbf{u}$ to $\mathbf{v}$ *without* being a dictatorship of the present —this seems reasonable, since the disparity between $\Phi(\mathbf{u})$ and $\Phi(\mathbf{v})$ is so large. Note also that, if $\Phi : \mathbb{U}^{\mathbb{N}} \longrightarrow \mathbb{R}$ is a (real-valued) IGWF, then the 'finiteness' condition is vacuously true, so that our definition of 'dictatorship' is then strictly weaker than Chichilnisky's. Thus, the property of 'nondictatorship' is stronger in our model than in hers.

Proposition 1 *Let $\mathbb{U} \subseteq \mathbb{R}$ be any nonsingleton set. Then ${}^*\sum$ is neither a weak dictatorship of the present, nor a weak dictatorship of the future on $\mathbb{U}^{\mathbb{N}}$.*

Proof: Let $x, y \in \mathbb{U}$ with $x < y$. Let $\mathbf{x} := (x, x, x, \ldots) \in \mathbb{U}^{\mathbb{N}}$ and $\mathbf{y} := (y, y, y, \ldots) \in \mathbb{U}^{\mathbb{N}}$.

To see nondictatorship of the present, let $\mathbf{u}, \mathbf{v} \in \mathbb{U}^{\mathbb{N}}$, and suppose ${}^*\sum \mathbf{u} \underset{*}{>} {}^*\sum \mathbf{v}$. We claim that ${}^*\sum$ does *not* myopically prefer $\mathbf{u}$ to $\mathbf{v}$. Indeed, for *any* $T \in \mathbb{N}$, we will show that ${}^*\sum(\mathbf{u}^T, \mathbf{x}_T) \underset{*}{<} {}^*\sum(\mathbf{v}^T, \mathbf{y}_T)$. To see this, let $U_N := \sum_{n=0}^N (\mathbf{u}^T, \mathbf{x}_T)_n$ and $V_N := \sum_{n=0}^N (\mathbf{v}^T, \mathbf{y}_T)_n$ for all $N \in \mathbb{N}$, and let $\mathbf{U} := (U_N)_{N=0}^\infty$ and $\mathbf{V} := (V_N)_{N=0}^\infty$; hence ${}^*\sum(\mathbf{u}^T, \mathbf{x}_T) = [\mathbf{U}]$ and ${}^*\sum(\mathbf{v}^T, \mathbf{y}_T) = [\mathbf{V}]$. Let $W := \sum_{t=0}^T (u_t - v_t)$, and $z := y - x > 0$. Then for all $N > T + W/z$, we have $V_N = U_N + (N-T)z - W > U_N$. Thus, $\{N \in \mathbb{N} \,;\, V_N > U_N\}$ is cofinite, hence in $\mathcal{F}$; thus $[\mathbf{V}] \underset{*}{>} [\mathbf{U}]$.

To see nondictatorship of the future, let $\mathbf{u}, \mathbf{v} \in \mathbb{U}^{\mathbb{N}}$. Suppose ${}^*\sum \mathbf{u} > {}^*\sum \mathbf{v}$, but ${}^*\sum \mathbf{u} - {}^*\sum \mathbf{v}$ is finite. We claim that ${}^*\sum$ does *not* eternally prefer $\mathbf{u}$ to $\mathbf{v}$.

If ${}^*\sum \mathbf{u} - {}^*\sum \mathbf{v}$ is finite, then there is some $B \in \mathbb{N}$ and $\mathbf{F} \in \mathcal{F}$ such that $-B < \sum_{t=0}^T u_t - \sum_{t=0}^T v_t < B$, $\forall\, T \in \mathbf{F}$. Thus $-2B < \sum_{t=T}^N u_t - \sum_{t=T}^N v_t < 2B$, $\forall\, T, N \in \mathbf{F}$. Thus, if $\mathbf{0} := (0, 0, 0, \ldots)$, then $-{}^*2B < {}^*\sum(\mathbf{0}^T, \mathbf{u}_T) - {}^*\sum(\mathbf{0}^T, \mathbf{v}_T) < {}^*2B$, $\forall\, T \in \mathbf{F}$. Let $x < y$ and $z := y - x > 0$ as before. If $T := 1 + \lceil 2B/z \rceil$, then $Tz > 2B$. Thus,

$$\begin{aligned} {}^*\sum(\mathbf{y}^T, \mathbf{v}_T) \underset{*}{-} {}^*\sum(\mathbf{x}^T, \mathbf{u}_T) &= {}^*\left(\sum_{t=0}^T (y - x)\right) \underset{*}{+} {}^*\sum(\mathbf{0}^T, \mathbf{v}_T) \underset{*}{-} {}^*\sum(\mathbf{0}^T, \mathbf{u}_T) \\ &\underset{*}{>} {}^*(Tz) \underset{*}{-} {}^*2B \quad \underset{*}{>} \quad {}^*2B \underset{*}{-} {}^*2B \quad = \quad {}^*0. \end{aligned}$$

Thus, ${}^*\sum(\mathbf{y}^T, \mathbf{v}_T) \underset{*}{>} {}^*\sum(\mathbf{x}^T, \mathbf{u}_T)$, so ${}^*\sum$ does *not* eternally prefer $\mathbf{u}$ to $\mathbf{v}$. □

Previous work. Nondiscounted, infinite-horizon, intergenerational social choice is similar to the problem of constructing a nondictatorial social choice function for a countably infinite population; this has been studied by Fishburn (1970), Candeal et al. (1992), Efimov and Koshevoy (1994), Lauwers (1993, 1997a,b), and others, and some solutions use ultrafilters (although none involve hyperreal numbers). For example, Chichilnisky and Heal (1997) constructed a social choice rule $\Psi : \mathbf{X}^{\mathbb{N}} \longrightarrow \mathbf{X}$ (where $\mathbf{X}$ was a topological space of preferences) that was continuous, respected unanimity, and nondictatorial (but not anonymous), by defining $\Psi(\mathbf{x})$ to be limit of the sequence $\mathbf{x} := (x_1, x_2, \ldots)$ along a free ultrafilter in $\mathbb{N}$. More generally, Kirman and Sondermann (1972) and Lauwers and Van Liedekerke (1995) have shown that every voting rule over a countable population is equivalent to an ultraproduct defined by some ultrafilter on $\mathbb{N}$. Campbell (1990) used ultrafilters to prove versions of Arrow's and Wilson's impossibility theorems for infinite-horizon intergenerational social choice.

Practicalities. To apply the IGWF ${}^*\sum : \mathbb{R}^{\mathbb{N}} \longrightarrow {}^*\mathbb{R}$, we need a concrete realization of ${}^*\mathbb{R}$, which requires an explicit specification of a free ultrafilter $\mathcal{F}$. Here we run into a practical problem: it is essentially impossible to explicitly specify a free ultrafilter; the proof of their existence is inextricably nonconstructive and uses the Axiom of Choice.[4] However, in many cases it may be sufficient to simply know that $\mathcal{F}$ exists.

Given two utility streams $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{\mathbb{N}}$, let $\mathbf{S} := \left\{T \in \mathbb{N} \,;\, \sum_{t=0}^{T} u_t > \sum_{t=0}^{T} v_t\right\}$; then ${}^*\sum \mathbf{u} \underset{*}{>} {}^*\sum \mathbf{v}$ if and only if $\mathbf{S} \in \mathcal{F}$. There are three possibilities: either $\mathbf{S}$ is finite, or $\mathbf{S}$ is cofinite (i.e. $\mathbf{S}^{\complement}$ is finite), or neither. Now, we know that $\mathcal{F}$ contains *all* cofinite sets, so if $\mathbf{S}$ is cofinite, then we know ${}^*\sum \mathbf{u} \underset{*}{>} {}^*\sum \mathbf{v}$, whereas if $\mathbf{S}$ is finite, then ${}^*\sum \mathbf{u} \underset{*}{<} {}^*\sum \mathbf{v}$. For example, suppose there are long-term equilibria $\overline{U}, \overline{V}$ such that $u_t \underset{t\to\infty}{\longrightarrow} \overline{U}$ and $v_t \underset{t\to\infty}{\longrightarrow} \overline{V}$ (as in §3 below). If $\overline{U} > \overline{V}$, then necessarily ${}^*\sum \mathbf{u} \underset{*}{>} {}^*\sum \mathbf{v}$, even if $\sum_{t=0}^{10000} u_t \ll \sum_{t=0}^{10000} v_t$. (For example, $\overline{V}$ might represent a future where short-term overconsumption leads to a long-term environmental catastrophe, whereas $\overline{U}$ represents a future where this catastrophe is averted). Likewise, if $\overline{U} < \overline{V}$, then ${}^*\sum \mathbf{u} \underset{*}{<} {}^*\sum \mathbf{v}$. If $\overline{U} = \overline{V}$, then we must look at short-term forecasts. For example, suppose there is some $T \in \mathbb{N}$ such that $\sum_{t=0}^{T} u_t > \sum_{t=0}^{T} v_t$ while $u_t = v_t$ for all $t \geq T$. Or suppose there exists $T \in \mathbb{N}$ and $\epsilon > 0$ such that $\sum_{t=0}^{T} u_t > \epsilon + \sum_{t=0}^{T} v_t$, while $\sum_{t=T+1}^{\infty} |u_t - v_t| < \epsilon$. In either case, clearly ${}^*\sum \mathbf{u} \underset{*}{>} {}^*\sum \mathbf{v}$.

But suppose that $\overline{U} = \overline{V}$ and neither $\mathbf{u}$ nor $\mathbf{v}$ clearly dominates the other. Or suppose the sequences $(u_t)_{t=0}^{\infty}$ and $(v_t)_{t=0}^{\infty}$ do *not* converge to long-term equilibria at all. Then $\mathbf{S}$ is neither finite nor cofinite; the situation is more ambiguous. Presumably we can only determine a finite fragment of $\mathbf{S}$; perhaps we can only estimate $\sum_{t=0}^{T} u_t$ and $\sum_{t=0}^{T} v_t$ for $T \leq 10^4$. We must guess whether $\mathbf{S} \in \mathcal{F}$, knowing only $\mathbf{S} \cap [1...10^4]$.

If $\mathbf{U} \subseteq \mathbb{N}$, its *lower Cesàro density* is defined $\underline{d}(\mathbf{U}) := \liminf_{T\to\infty} |\mathbf{U} \cap [1...T]|/T$; clearly $0 \leq \underline{d}(\mathbf{U}) \leq 1$. If $\mathcal{D} := \{\mathbf{U} \subseteq \mathbb{N} \,;\, \underline{d}(\mathbf{U}) = 1\}$, then there is a free ultrafilter which contains $\mathcal{D}$ (this follows from Zorn's Lemma, because $\mathcal{D}$ itself is a free filter).

[4] A similar problem afflicts the IGWF χ in eqn. (1); specifying a Hahn-Banach extension of lim is equivalent to specifying a free ultrafilter. See also Chichilnisky and Heal (1997).

So, assume without loss of generality that $\mathcal{F} \supset \mathcal{D}$. Thus, if $|\mathbf{S} \cap [1...10^4]| / 10^4 \approx 1$, then this suggests that $\underline{d}(\mathbf{S}) = 1$, which implies that $\mathbf{S} \in \mathcal{F}$, and hence ${}^*\sum \mathbf{u} \underset{*}{>} {}^*\sum \mathbf{v}$.

There are two problems with this approach: (i) There is no lower bound on the speed with which $|\mathbf{S} \cap [1...T]|/T$ converges to $\underline{d}(\mathbf{S})$; hence the fact that $|\mathbf{S} \cap [1...10^4]| / 10^4 \approx 1$ proves nothing. (ii) What if $|\mathbf{S} \cap [1...10^4]| / 10^4 \approx 0.5$? What then? In this case, our forecasts are sufficiently ambiguous that the best response is probably to be indifferent between $\mathbf{u}$ and $\mathbf{v}$. Presumably, in most situations, the contrast between forecasts will be starker, and our choice will be clearer.

2 Discounting without time preferences

To make intertemporal choices under uncertainty, an agent compares different scenarios which generate different time-sequences of future expected utility. This future utility is determined by the future values of 'instrumental' variables such as consumption or income. We say the agent exhibits *pure time preferences* if the agent discounts future expected utility itself (e.g. due to 'impatience' or 'myopia'). However, even *without* pure time preferences, the agent may discount instrumental variables, because their values are increasing over time (e.g. due to economic growth) while offering diminishing marginal utility (e.g. because utility is a concave function of income). We call this *instrumental discounting*; we will mathematically model it in §4.

For example, suppose $U(y) = \log(y)$, where U is utility and y is income. Then $U'(y) = 1/y$. Thus, if I expect my income in 2038 to be double my income in 2008, then each dollar of additional income in 2038 yields *half* the marginal utility of an additional dollar in 2008; hence in 2008 I will discount 2038 marginal income by 50%, even if do not discount 2038 utility itself at all. (However, if I expect to be *poorer* in 2038, then I will actually value 2038 income *more* than 2008 income).

An agent may also discount the future due to uncertainty. If future utility depends on the future value of instrumental variables, then intertemporal decisions require predictions about these variables. These predictions are always uncertain, and this uncertainty increases as the prediction date moves further into the future. Thus, the expected utility in the far future converges to some equilibrium value which is often more or less independent of the details of short-term decisions. Thus, even *without* pure time preferences or instrumental discounting, the agent's intertemporal decisions may seem to 'discount' the long-term consequences of her actions —not because the agent doesn't *care* about these consequences, but simply because the agent has no way of predicting (and hence optimizing) these consequences. We call this *informational discounting*; we will mathematically model it in §3.

Cowen and Parfit (1992) and others have argued that a social planner should have *no* pure time-preferences: the utility of future people is just as important as the utility of present people. In §1 we constructed an IGWF ${}^*\sum$ with this property. Arguably, individuals should also have no pure time preferences: your future utility is just as important as your present utility. It is true that people exhibit a nonzero discount rate (e.g. they borrow money at positive real interest rates); furthermore, many models of intertemporal choice require discounting for equilibria to exist.[5] But this can be explained as informational or instrumental discounting, without any pure time preferences.

3 Informational discounting due to burgeoning uncertainty

Suppose the agent's utility is described by a function $\Upsilon : \mathcal{A} \times \mathbb{N} \times \mathbf{X} \longrightarrow \mathbb{R}$, where $\mathcal{A}$ is a set of 'strategies' (controlled by the agent) and $\mathbf{X}$ is a set of exogenous 'world states' (not controlled by the agent), and where $\Upsilon_t^a(x) := \Upsilon(a, t, x)$ is the utility obtained from state x at time t, assuming strategy a was chosen at time zero.[6]

Suppose that the world-state changes over time according to a Markov process. Formally, let $\mathcal{P}(\mathbf{X})$ be the space of all probability measures over $\mathcal{X}$, and let $\Phi : \mathcal{P}(\mathbf{X}) \longrightarrow \mathcal{P}(\mathbf{X})$ be a linear transformation (the *transition probability operator*), so that, if $\rho_t \in \mathcal{P}(\mathbf{X})$ is the probability distribution of an unknown world-state at time t, then $\rho_{t+1} := \Phi(\rho_t)$ will be the resulting probability distribution of the unknown world-state at time $t + 1$. In particular, if the state of the system at time $t = 0$ is known to be $x \in \mathcal{X}$, then the probability distribution at time $t = 1$ is $\Phi(\delta_x)$ (where δ_x is the pointmass at x), and the distribution at time $t = 2$ is $\Phi^2(\delta_x)$, and so on.

A probability measure $\eta \in \mathcal{P}(\mathbf{X})$ is *stationary* if $\Phi(\eta) = \eta$ —that is, η is an eigenvector of Φ with eigenvalue 1. Under fairly general conditions, Φ admits a unique stationary measure η, which is a globally attracting fixed point for the action of Φ on $\mathcal{P}(\mathbf{X})$; this is the *Perron-Frobenius Theorem.*[7] Let $\mathrm{Spec}(\Phi) \subset \mathbb{C}$ be the set of all eigenvalues of Φ, and let $\lambda := \sup\{|c| \; ; \; c \in \mathrm{Spec}(\Phi) \text{ and } c \neq 1\}$. If $\lambda < 1$, then for any $\rho_0 \in \mathcal{P}(\mathbf{X})$, if $\rho_t := \Phi^t(\rho_0)$ for all $t \in \mathbb{N}$, then we have

$$\|\rho_t - \eta\|_1 \quad \leq \quad \lambda^t \, \|\rho_0 - \eta\|_1 \quad \underset{t\to\infty}{\longrightarrow} \quad 0, \tag{3}$$

where $\|\bullet\|_1$ is the total variation norm on $\mathcal{P}(\mathbf{X})$. In other words, the sequence $\{\rho_t\}_{t=0}^{\infty}$ converges to η exponentially.

Let $\widetilde{\mathbf{x}} := (\widetilde{x}_t)_{t=0}^{\infty} \in \mathbf{X}^{\mathbb{N}}$ be a random path from this Markov process. Any strategy $a \in \mathcal{A}$ determines a random utility stream $\widetilde{\mathbf{u}}^a := (\widetilde{u}_t^a)_{t=0}^{\infty} \in \mathbb{R}^{\mathbb{N}}$, where $\widetilde{u}_t^a := \Upsilon_t^a(\widetilde{x}_t)$. If ρ_t is the distribution of $\widetilde{x}_t$, then the expected value of $\widetilde{u}_t$ is given by

$$\overline{u}_t^a \quad := \quad \int_{\mathbf{X}} \Upsilon_t^a(x) \; d\rho_t[x].$$

[5] See e.g. Araujo (1985) for the 'need for impatience' in intertemporal exchange economies, or Muthoo (1999) for the role of discounting in 'alternating offers' bargaining models.

[6] The agent commits to a at time zero, but she executes a over time; thus a could include contingent clauses like 'If the state becomes x at time t, then do the following at time $t+1$'. However, it is also possible that the actions dictated by a before time t will constrain the agent's options at time $t+1$, e.g. due to sunk costs, contractual commitments, etc.

[7] See e.g (Lind and Marcus, 1995, Thm 4.2.3) or (Borkar, 1995, Thm. 5.3.2).

Let $\overline{U}_t^a := \int_{\mathbf{X}} \Upsilon_t^a(x)\ d\eta[x]$ and let $M_t^a := \|\Upsilon_t^a\|_\infty$, and assume that the sequence $\{M_t^a\}_{t=0}^\infty$ is bounded or grows at most subexponentially. Then

$$\begin{aligned} |\overline{u}_t^a - \overline{U}_t^a| &= \left| \int_{\mathbf{X}} \Upsilon_t^a(x)\ d(\rho_t - \eta)[x] \right| \quad \leq \quad M_t^a \|\rho_t - \eta\|_1 \\ &\underset{(3)}{\leq} M_t^a\, \lambda^t\, \|\rho_0 - \eta\|_1 \quad \xrightarrow[t\to\infty]{} \quad 0 \qquad \text{(exponentially).} \end{aligned} \tag{4}$$

For any $a, b \in \mathcal{A}$, if $\liminf_{t\to\infty}(\overline{U}_t^a - \overline{U}_t^b) > 0$, then a *asymptotically dominates* b: without any knowledge of the future except for η, we can see that a is a better long-term strategy. An infinitely patient agent would never choose an asymptotically dominated strategy; hence we can assume without loss of generality that all asymptotically dominated strategies have already been eliminated from $\mathcal{A}$. Thus, for all $a, b \in \mathcal{A}$, we have $\lim_{t\to\infty} |\overline{U}_t^a - \overline{U}_t^b| = 0$ —neither a nor b asymptotically dominates the other. Combining this with eqn.(4), we conclude that $|\overline{u}_t^a - \overline{u}_t^b| \xrightarrow[t\to\infty]{} 0$. (Indeed, if $\overline{U}_t^a = 0 = \overline{U}_t^b$ for all $a, b \in \mathcal{A}$ and $t \in \mathbb{N}$, as in the two examples below, then $|\overline{u}_t^a - \overline{u}_t^b| \xrightarrow[t\to\infty]{} 0$ exponentially). This means: given only information about the present (encoded in ρ_0), the choice between a and b will be decided mainly by the *short-term* behaviour of the expected utility streams $(\overline{u}_t^a)_{t=0}^\infty$ and $(\overline{u}_t^b)_{t=0}^\infty$; hence the agent behaves *as if* she discounts the long-term future. We will illustrate this with two examples.

3.1 Mortality. Let $\mathbf{X} := \{0, 1\}$, where 1 represents 'alive' and 0 represents 'dead'. Thus, $\mathcal{P}(\mathbf{X}) := \{(p_0, p_1) \in \mathbb{R}^2_{\not-}\ ;\ p_0 + p_1 = 1\}$. Let $\Phi : \mathcal{P}(\mathbf{X}) \longrightarrow \mathcal{P}(\mathbf{X})$ be the linear transformation with matrix $\left[\begin{smallmatrix} 1 & d \\ 0 & 1-d \end{smallmatrix}\right]$. In other words, a living person has a probability d of dying during each period; a dead person stays dead. This *Poisson process* roughly describes the problem faced by a mortal human who cannot predict her own mortality. The unique stationary probability measure is $\eta = \delta_0$, and the maximal nonunit eigenvalue is $\lambda = 1 - d$. Suppose that, for all $a \in \mathcal{A}$, $\Upsilon_t^a(0) = 0$ and $\Upsilon_t^a(1) = u_t^a > 0$. Then, having chosen strategy a at time zero, the expected utility at time t is $\overline{u}_t^a = \lambda^t u_t^a$. Thus, the expected lifetime utility resulting from strategy a is ${}^*\sum_{t=0}^\infty \lambda^t u_t^a$ —the traditional exponentially discounted sum of future utility.

For example, suppose $a, b \in \mathcal{A}$ are two 'investment strategies'. Assume a represents 'unsustainable immediate gratification', so that $(u_t^a)_{t=0}^\infty$ is a sequence decreasing to zero, whereas b represents 'profitable long-term investment', so that $(u_t^b)_{t=0}^\infty$ is a sequence increasing to infinity. However, suppose that b also requires some short-term sacrifice, so that $u_t^b < u_t^a$ for all $t \in [1...T]$. Then it is quite easy to construct examples where $\sum_{t=0}^\infty \lambda^t u_t^b < \sum_{t=0}^\infty \lambda^t u_t^a$ —in other words, a mortal might rationally choose immediate gratification over long-term investment.

Besides mortal humans, the Poisson process is also relevant to intertemporal choice by firms, for two reasons. (1) Firms exist to generate dividends for human shareholders, who (being mortal) want the dividends now, rather than later. (2) A firm itself is 'mortal': it might go bankrupt (with little warning), at which point its assets will be liquidated to pay creditors, and the shareholders will likely get nothing. Because of this risk, shareholders again prefer dividends now, rather than later.

The Poisson process can also induce exponential discounting in other intertemporal economic activities susceptible to sudden, exogenous termination. For example, in 'alternating offers' models of bargaining (Muthoo, 1999), the state 0 might represent the sudden termination of negotiations because the counterparty has received a better 'outside offer'. For a social planner, the state 0 might represent apocalypse.

3.2 Mean Reversion. Let $\mathbf{X} := \mathbb{R}^n$, where each coordinate represents some economically relevant variable (e.g. weather conditions, commodity prices, etc.). An $\mathbb{R}^n$-valued Markov process is ***mean-reverting*** if it has a unique, exponentially attracting stationary probability measure $\eta \in \mathcal{P}(\mathbb{R}^n)$ with a finite mean. After a change of coordinates we can assume this mean is $\mathbf{0}$. (The most familiar example is the *Ornstein-Uhlenbeck* process, a random walk which exponentially 'tries to converge' to $\mathbf{0}$, while being constantly perturbed by Gaussian random noise. In this case, η is a multivariate normal distribution with mean $\mathbf{0}$.)

If $(\widetilde{\mathbf{x}}_t)_{t=0}^{\infty} \in (\mathbb{R}^n)^{\mathbb{N}}$ is a random path from a mean-reverting process, and $\overline{\mathbf{x}}_t \in \mathbb{R}^n$ is the conditional expectation of $\widetilde{\mathbf{x}}_t$ given knowledge of $\widetilde{\mathbf{x}}_0$, then $\lim_{t\to\infty} \overline{\mathbf{x}}_t = \mathbf{0}$, independent of the value of $\widetilde{\mathbf{x}}_0$. For any $a \in \mathcal{A}$ and $t \in \mathbb{N}$, suppose that $\Upsilon_t^a : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a linear utility function, with spectral radius bounded by some $M < \infty$. Then $\overline{u}_t^a := \Upsilon_t^a(\overline{\mathbf{x}}_t) \xrightarrow[t\to\infty]{} 0$ for all $a \in \mathcal{A}$. That is: for any strategy and any initial condition, the long-term expected utility is zero.

For example, suppose a is a business strategy and Υ is profit. If $\overline{u}_t^a \xrightarrow[t\to\infty]{} 0$, then a might generate positive expected profits in the short term, but it will converge to zero expected profits in the long term. This describes innovation-driven profit in a perfectly competitive market with no barriers to entry. In the short term, the 'innovation' a yields positive profit because the firm can capture monopoly rents. But eventually, imitators enter the market, and competition drives profits down to zero.

4 Intergenerational choice with a growing economy and population

Let W represent society's endowment of resources at time zero, which can either be consumed or invested in production. Assume an exogenous 'yield' rate $\gamma \geq 1$ for any investment. Thus, if unconsumed, the endowment W will grow to size γW at time 1, and to size $\gamma^t W$ at time t. Thus, γ is the maximum growth rate of the economy; in reality, the economy will grow more slowly, because some yield will be consumed, not reinvested. A consumption stream $\mathbf{c} = (c_t)_{t=0}^{\infty} \in \mathbb{R}_{\not-}^{\mathbb{N}}$ is feasible if and only if

$$\sum_{t=0}^{\infty} \gamma^{-t} c_t \quad \leq \quad W. \tag{5}$$

Let $\pi \geq 1$ be an exogenous population growth rate, and assume that the population at time 0 is 1, so that the population at time t is π^t (presumably $\pi \leq \gamma$; otherwise we have a Malthusian scenario). Let $\alpha \in (-\infty, 1]$, and assume that per capita utility is determined by per capita consumption x via the function u_α defined: $u_\alpha(x) := x^\alpha/\alpha$ if $\alpha \neq 0$ and $u_\alpha(x) := \log(x)$ if $\alpha = 0$. Thus, u_α is a concave, increasing function, and $u'_\alpha(x) = x^{\alpha-1}$, for any $\alpha \in (-\infty, 1]$. Thus, if c_t is the aggregate consumption at time t, then the per capita consumption is c_t/π^t, so the per capita utility is $u_\alpha(c_t/\pi^t)$, so the aggregate utility is $\pi^t u_\alpha(c_t/\pi^t)$.

Let $\delta \in (0, 1]$ be a discount factor. Thus, $\delta < 1$ if the social planner has pure time preferences, or resorts to informational discounting as in §3.1, whereas $\delta = 1$ if she has no time preferences and perfect foreknowledge. The consumption stream $\mathbf{c} = (c_t)_{t=0}^\infty$, generates a (hyperreal) total future aggregate utility of

$$U(\mathbf{c}) \quad := \quad {}^*\sum_{t=0}^{\infty} \delta^t \pi^t u_\alpha(c_t/\pi^t) \ \in \ {}^*\mathbb{R}, \tag{6}$$

where ${}^*\sum$ is defined as in §1. Thus, for any $t \in \mathbb{N}$, the marginal utility of c_t is given:

$$\partial_t U(\mathbf{c}) \quad = \quad \delta^t \pi^t u'_\alpha(c_t/\pi^t)/\pi^t \quad = \quad \delta^t \pi^{t(1-\alpha)} c_t^{\alpha-1}. \tag{7}$$

(Even if the sum (6) is hyperfinite, the derivative (7) is finite, so long as $c_t > 0$). If the consumption stream $\mathbf{c}$ optimizes U with respect to the budget constraint (5), then there exists $\lambda > 0$ such that for all $t \in \mathbb{N}$, we have $\partial_t U(\mathbf{c}) = \lambda\gamma^{-t}$. Setting $t = 0$ and substituting the expression (7) we get $c_0^{\alpha-1} = \lambda$. Thus, $\mathbf{c}$ satisfies:

$$\delta^t \pi^{t(1-\alpha)} c_t^{\alpha-1} \quad = \quad \gamma^{-t} c_0^{\alpha-1}, \qquad \forall\, t \in \mathbb{N}. \tag{8}$$

Let $\theta := \pi(\delta\gamma)^{\frac{1}{1-\alpha}}$. Then simplifying (8) yields $$c_t \quad = \quad \theta^t c_0, \qquad \forall\, t \in \mathbb{N}. \tag{9}$$

If $\mathbf{c}$ is optimal, then (5) is an equality. Substituting (9) into (5), we get

$$W \quad = \quad \sum_{t=0}^{\infty} \gamma^{-t} c_t \quad = \quad \sum_{t=0}^{\infty} (\theta/\gamma)^t c_0 \underset{(*)}{=} \frac{c_0}{1-(\theta/\gamma)}, \tag{10}$$

$$\text{where } (*) \text{ is true} \quad \iff \quad \Big(\theta/\gamma \ < \ 1\Big) \underset{(\dagger)}{\iff} \left(\delta \ < \ \frac{\gamma^{-\alpha}}{\pi^{1-\alpha}}\right). \tag{11}$$

(here (†) is because $1 - \alpha > 0$). If (11) holds, then we deduce $c_0 = [1 - (\theta/\gamma)]W$; we can then substitute this into (9) to compute the optimal consumption path.

For example, suppose $\alpha = -1$ and $\delta = 1$; then condition (11) is equivalent to $\gamma > \pi^2$, while $\theta = \pi\sqrt{\gamma}$. Suppose population grows at 3% per year, while investment yields 8.16%; then we have $\pi = 1.03$ and $\gamma = 1.0816 = (1.04)^2 > \pi^2$, so (11) is satisfied, with $\theta = (1.03)(1.04) = 1.0712$. Thus, $c_0 = [1 - (\theta/\gamma)]W \approx 0.0096W$, and $c_t = (1.0712)^t \cdot c_0$. Thus, aggregate consumption (i.e. GDP) grows at 7.12% per year, while per capita consumption grows at 4% per year.

Note that the social planner does *not* aim for the maximum possible economic growth rate of 8.16%. She sacrifices some future growth for present consumption, because she 'instrumentally discounts' future prosperity, as described in §2.

If condition (11) does not hold, then the sum (10) is infinite, for any $c_0 > 0$. This means that the optimization problem has no nontrivial solution. Intuitively, this is because, for any $t \in \mathbb{N}$ with $c_t > 0$, there is always some $T > t$ such that c_T is small enough that one can increase the value of U by decreasing c_t to 0 and increasing c_T to $c_T + \gamma^{T-t} c_t$ —i.e. by 'deferring' gratification from time t to time T. Inductively, one ends up deferring all gratification until eternity, so that $c_t = 0$ for all $t \in \mathbb{N}$.

If $\delta = 1$, then this 'Paradox of Eternal Deferral' can occur in two opposite ways. If $\alpha < 0$, then the Paradox occurs only if $\gamma \leq \pi^{1-\frac{1}{\alpha}}$ —i.e. if yield is too small, relative to population growth. However, if $\alpha \geq 0$, then the Paradox occurs for *any* $\pi \geq 1$ and $\gamma > 1$. In other words, if utility functions have the form $u(x) = \log(x)$ or $u(x) = x^\alpha$ for $0 < \alpha \leq 1$, then nondiscounted utilitarian intergenerational social choice is *impossible* in a scenario of exponential economic growth.

There are three ways to resolve the Paradox. (1) Insist that $\alpha < 0$ (an assumption about human psychology with no *a priori* justification) and hope that γ is large enough. (2) Insist that $\delta < 1$ —i.e. either that the social planner has pure time preferences (contradicting the rationale of this entire article) or at least, that she 'informationally discounts' the future as in §3.1. (3) Reject the assumption of exponential economic growth. The assumption of a constant 'yield' $\gamma > 1$ for investment is equivalent to a technology with constant returns to scale. If instead we assume decreasing returns to scale, then economic growth will slow down and eventually stop (and hopefully, population growth along with it). This seems plausible if we imagine a society confined to a finite resource base (e.g. Planet Earth). However, this is beyond the scope of the very simple model we have presented here.

References

Anderson, R. M., 1991. Nonstandard analysis with applications to economics. In: Handbook of mathematical economics, Vol. IV. North-Holland, Amsterdam, pp. 2145–2208.

Araujo, A., 1985. Lack of Pareto optimal allocations in economies with infinitely many commodities: the need for impatience. Econometrica 53 (2), 455–461.

Arkeryd, L. O., Cutland, N. J., Henson, C. W. (Eds.), 1997. Nonstandard analysis. Vol. 493 of NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences. Kluwer Academic Publishers Group, Dordrecht, theory and applications.

Bewley, T., 1970. Equilibrium theory with an infinite-dimensional commodity space. Ph.D. thesis, UC Berkely.

Borkar, V. S., 1995. Probability theory. Universitext. Springer-Verlag, New York, an advanced course.

Brown, D. J., Lewis, L. M., 1981. Myopic economic agents. Econometrica 49 (2), 359–368.

Campbell, D. E., 1990. Intergenerational social choice without the Pareto principle. J. Econom. Theory 50 (2), 414–423.

Candeal, J. C., Induráin, E., Uriarte, J. R., 1992. Some issues related to the topological aggregation of preferences. Soc. Choice Welf. 9 (3), 213–227.

Caplin, A., Leahy, J., 2004. The social discount rate. Journal of Political Economy 112 (6), 1257–1268.

Chichilnisky, G., 1996. An axiomatic approach to sustainable development. Soc. Choice Welf. 13 (2), 231–257.

Chichilnisky, G., Heal, G., 1997. Social choice with infinite populations: construction of a rule and impossibility results. Soc. Choice Welf. 14 (2), 303–318, topological social choice.

Cowen, T., Parfit, D., 1992. Against the social discount rate. In: Laslett, P., Fishkin, J. S. (Eds.), Justice between age groups and generations. Philosophy, Politics, and Society. Yale UP, New Haven, CT, pp. 144–61.

Efimov, B. A., Koshevoy, G. A., 1994. A topological approach to social choice with infinite populations. Math. Social Sci. 27 (2), 145–157.

Fishburn, P. C., 1970. Arrow's impossibility theorem: concise proof and infinite voters. J. Econom. Theory 2, 103–106.

Kirman, A. P., Sondermann, D., 1972. Arrow's theorem, many agents, and invisible dictators. J. Econom. Theory 5 (2), 267–277.

Lauwers, L., 1993. Infinite Chichilnisky rules. Econom. Lett. 42 (4), 349–352.

Lauwers, L., 1997a. A note on weak ∞-Chichilnisky rules. Soc. Choice Welf. 14 (2), 357–358, topological social choice.

Lauwers, L., 1997b. Topological aggregation, the case of an infinite population. Soc. Choice Welf. 14 (2), 319–332, topological social choice.

Lauwers, L., Van Liedekerke, L., 1995. Ultraproducts and aggregation. J. Math. Econom. 24 (3), 217–237.

Lind, D., Marcus, B., 1995. An introduction to symbolic dynamics and coding. Cambridge UP, New York.

Muthoo, A., 1999. Bargaining theory with applications. Cambridge UP, Cambridge, UK.

Rashid, S., 1987. Economies with many agents: An approach using nonstandard analysis. Johns Hopkins University Press, Baltimore, MD.

von Weizacker, January 1967. Lemmas for a theory of approximate optimal growth. Review of Economic Studies 34.1 (97), 143–151.