Centre for Climate Change Economics and Policy

An ESRC Research Centre

Grantham Research Institute on Climate Change and the Environment

Heterogeneous intergenerational altruism

Antony Millner

February 2016

Centre for Climate Change Economics and Policy
Working Paper No. 256

Grantham Research Institute on Climate Change and the Environment
Working Paper No. 226

ESRC Economic & Social Research Council · The Grantham Foundation for the Protection of the Environment · University of Leeds · LSE The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council. Its second phase started in 2013 and there are five integrated research themes:

1. Understanding green growth and climate-compatible development
2. Advancing climate finance and investment
3. Evaluating the performance of climate policies
4. Managing climate risks and uncertainties and strengthening climate services
5. Enabling rapid transitions in mitigation and adaptation

More information about the Centre for Climate Change Economics and Policy can be found at: http://www.cccep.ac.uk.

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and the Global Green Growth Institute. It has nine research programmes:

1. Adaptation and development
2. Carbon trading and finance
3. Ecosystems, resources and the natural environment
4. Energy, technology and trade
5. Future generations and social justice
6. Growth and the economy
7. International environmental negotiations
8. Modelling and decision making
9. Private sector adaptation, risk and insurance

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: http://www.lse.ac.uk/grantham.

Heterogeneous intergenerational altruism

Antony Millner*[1]

[1]London School of Economics and Political Science

February 11, 2016

Abstract

Agents exhibit pure intergenerational altruism if they care not just about the consumption utility experienced by future generations, but about their total wellbeing. If all generations are altruistic, each generation's wellbeing depends on the wellbeing of its descendants. Thus pure intergenerational altruism causes generations' preferences to be interdependent. While existing models study the relationship between pure intergenerational altruism and conventional time preferences, they assume that altruistic preferences are homogeneous across society. In effect, agents impose their own preferences on future generations, whether they share them or not. By contrast, we study pure intergenerational altruism when agents' preferences are heterogeneous and fully non-paternalistic, i.e. they evaluate the wellbeing of future agents according to their own sovereign intergenerational preferences. We demonstrate that homogeneous models of intergenerational altruism over (under) estimate the weight an agent places on future utilities if she is less (more) altruistic than average. Moreover, all non-paternalistic agents agree on the appropriate long-run utility discount rate, regardless of their preferences. In general, existing derivations of exponential or quasi-hyperbolic time preferences from homogeneous models of pure intergenerational altruism are not robust to heterogeneity.

Keywords: Time preferences, heterogeneity, intergenerational altruism, interdependent preferences, quasi-hyperbolic discounting

*London School of Economics and Political Science, Houghton St, London, WC2A 2AE, UK (email: a.millner@lse.ac.uk). I acknowledge support from the Centre for Climate Change Economics and Policy, funded by the UK Economic and Social Research Council.

1 Introduction

An agent exhibits Pure Intergenerational Altruism (PIA) if she cares not only for the utility that a future generation experiences due to its own consumption, but also for that generation's concern for the wellbeing of its descendants. For example, a mother who cares about the wellbeing of her children, and recognizes that her children's wellbeing depends not only on how their own lives go, but also on how *their* children's (i.e. her grandchildren's) lives go, exhibits a degree of PIA. A mother who neglects the impact of her grandchildren's lives on her own children's wellbeing is an intergenerational paternalist – her children's preferences enter her own only in a limited way. Just as parents who neglect their children's intergenerational concerns act paternalistically towards them, so too do ethical observers who feel altruistic towards future societies, but neglect those societies' own altruistic feelings when evaluating their wellbeing. Moreover, since members of those future societies will invariably have ethical disagreements about how to value the wellbeing of their descendants, even an agent who does account for the altruistic feelings of future societies acts paternalistically if she assumes that all members of those societies will share her own preferences. This paper studies PIA preferences when agents are fully non-paternalistic – they account both for the altruistic feelings of future generations towards their descendants, and also recognize that not all members of future societies will share their own preferences. We show that preference heterogeneity qualitatively alters existing results on the relationship between PIA and conventional time preferences, and that non-paternalism forces all agents to agree on long-run utility discount rates, regardless of their preferences.

The preferences of agents who exhibit PIA are defined in terms of the preferences of their descendants, which are themselves defined in terms of the preferences of *their* descendants, ad infinitum. In order to be operationalized these recursively defined preference systems must be disentangled into equivalent paternalistic preferences over dynamic streams of utility. The simplest example of this is what we will refer to as first order PIA. In this case an agent in generation τ cares about its own generation's utility U_τ, and the (discounted) wellbeing of the next generation. Denoting the wellbeing of generation τ by V_τ, we have

$$V_\tau = U_\tau + \beta V_{\tau+1}, \tag{1}$$

where $\beta \in [0, 1)$ is a discount factor. We say that these preferences exhibit first order PIA since they depend only on the wellbeing of the first generation following generation

τ. Repeated substitution of this expression into itself shows that these preferences are equivalent to paternalistic preferences over infinite utility streams $(U_\tau, U_{\tau+1}, U_{\tau+2}, \ldots)$ of the form:

$$V_\tau = \sum_{s=0}^{\infty} \beta^s U_{\tau+s}. \tag{2}$$

This is perhaps the simplest justification for the use of discounted utilitarian time preferences in intergenerational decision-making.[1] This example illustrates how PIA preferences can be translated into more familiar paternalistic time preferences. The literature has focussed on how the properties of paternalistic time preferences relate to underlying PIA preferences, and we will do the same.

As a theory of pure intergenerational altruism, first order PIA is limited in two important respects. The first limitation is that the altruistic feelings of the current generation extend only to their immediate descendants. In general, we might expect altruistic agents to care not just about the wellbeing of their children's generation, but also about their grandchildren's generation, their great-grandchildren's generation, and so on. Models of higher order PIA have been extensively studied – we review these contributions in Section 1.1 below. A second limitation – the assumption that intergenerational preferences are homogeneous – has however not been investigated before.[2]

A common feature of ethical judgements such as how much weight to place on the well-being of future generations is that, even if all individuals agree to subject their views to rigorous scrutiny in open debate, invariably no universally agreed consensus emerges (Sen, 2010). We may be able to agree to rule certain preferences out on ethical grounds, but this does not imply that this process will terminate with all but one preference being excluded. For example, a devotee of agent-relative ethics may believe that the present generation is justified in treating itself as more important than future generations, causing her to discount the next generation's wellbeing more heavily than all generations after that (Arrow, 1999). Another observer may believe that such a stance is ethically indefensible, and advocate an exponential discounting rule. In general, ethical agents may have good-faith

[1] Of course, discounted utilitarianism has other desirable properties. It can be derived from primitive axioms on preferences over consumption streams (e.g. Koopmans, 1960; Fishburn and Rubinstein, 1982), by imposing time consistency on generalized separable and time invariant preferences (Strotz, 1955), or by assuming that individuals/humanity faces a constant probability of death/extinction (Yaari, 1965; Dasgupta and Heal, 1979). None of these derivations however accounts for PIA, which has independent ethical appeal.

[2] Some authors, e.g. Bergstrom (1999); Pearce (2008), do study systems of interdependent preferences in which *static* other-regarding preferences are heterogeneous across individuals. We clarify how our model extends their work in Section 1.1.

disagreements about both the functional form of discounting schedules, and the appropriate values of numerical parameters in a given functional specification. This contention is supported by diverse evidence, from persistent disputes between leading economists about the appropriate pure rate of time preference for the analysis of climate policy (e.g. Nordhaus, 2008; Stern, 2007; Weitzman, 2007), to explicit surveys of economists' normative prescriptions for intergenerational discount rates, which show substantial variation (Drupp et al., 2014).

If intergenerational preferences are heterogeneous, homogeneous models of PIA such as that in (1) are still subject to a paternalist critique. This occurs since an ethical agent with homogeneous PIA preferences neglects the fact that future societies' preferences may not coincide with her own. This amounts to a 'mother knows best' attitude, in which individuals in the current generation evaluate the wellbeing achieved by their descendants using their own intergenerational preferences, whether they subscribe to them or not. This is a fundamental limitation, since if intergenerational preferences are in fact heterogeneous, homogeneous PIA models are guilty of precisely the kind of paternalism they were designed to combat.

To address this limitation we generalize existing homogeneous models of PIA to account for heterogeneous preferences. The agents in our model are fully non-paternalistic, in that they evaluate the wellbeing of those who disagree with them using their own sovereign preferences. Despite this, all agents still have complete freedom to specify their intergenerational preferences as they wish – non-paternalism requires agents to be sensitive to other agents' definitions of their own wellbeing, but does not determine how much weight any given agent should place on future societies. We use our model to investigate three questions: how are agents' equivalent paternalistic time preferences affected by the presence of heterogeneity, how do heterogeneous agents' preferences aggregate into social preferences, and how do different agents' time preferences behave in the long run?

We begin by studying these questions in a first order PIA model. When individuals in the current generation only account for the heterogeneous preferences of their immediate descendants, their equivalent paternalistic time preferences are no longer exponential as suggested by (1–2), but rather correspond to the quasi-hyperbolic '$\beta\delta$' preferences of Phelps and Pollak (1968), where the long-run discount factor δ is common to all individuals.[3] However, while every individual in the population has quasi-hyperbolic preferences, social

[3]Homogeneous PIA models can also be used to derive quasi-hyperbolic time preferences, but require intergenerational preferences to be exponential, i.e. of infinite order. See Saez-Marti and Weibull (2005); Galperti and Strulovici (2015), and our discussion in Section 3.

time preferences are exponential. First order PIA with heterogeneous preferences thus gives rise to a population of present-biased agents, and simultaneously generates a time consistent welfare criterion that is derived directly from agents' preferences. This is a unique feature of this model.

We then generalize to models of PIA of arbitrary, possibly infinite, order. We show that if agents' intergenerational preferences can be ordered there is a systematic bias in homogeneous models of PIA. The intergenerational discount factors of those agents who are more altruistic than average move down relative to the equivalent homogeneous model, and vice versa for those agents who are less altruistic than average. The presence of heterogeneity thus causes agents' equivalent paternalistic time preferences to 'mean revert'. This convergence dynamic plays out more dramatically at the level of discount rates. We prove that if intergenerational preferences are heterogeneous and non-paternalistic, then all agents should agree on the appropriate long-run discount rate on future utilities, no matter what their intergenerational preferences are. This strong convergence result is a direct consequence of non-paternalism and the intergenerational entanglement of agents' preferences. Finally, we apply our model to a recent axiomatization of PIA preferences by Galperti and Strulovici (2015). We show that the derivation of quasi-hyperbolic time preferences that falls out of their preference representation is not robust to the presence of heterogeneity. If preferences are heterogeneous, non-paternalistic, and obey their axioms, neither individuals nor society can have quasi-hyperbolic time preferences. Thus, the presence of heterogeneity qualitatively alters the findings of homogenous models of PIA.

The paper is structured as follows. We discuss related literature next, before presenting the model and our main results in Section 2. This section presents the general model, discusses the case of first order heterogeneous PIA to build intuition, and then derives general results on the mean reverting effect of heterogeneity, the asymptotic convergence of all agents' discount rates, and conditions on the primitives of the model that ensure that agents' equivalent paternalistic time preferences are well behaved. Section 3 applies these results to the axiomatization of PIA preferences in Galperti and Strulovici (2015), and Section 4 concludes.

1.1 Related literature

This paper brings together two strands of literature. The first line of work deals with pure intergenerational altruism, while the second deals with heterogeneous time preferences and their aggregation.

The possibility that individuals' wellbeing could depend on the wellbeing of others is an old idea in economic theory; it is discussed by e.g. Pigou (1903), Fisher (1926), and Becker (1974). Pure intergenerational altruism first appeared in Barro (1974), who studies first order homogeneous PIA preferences. Barro and Becker (1989) apply these preferences in a model of fertility choice. Extensions of the first order PIA model to higher order altruism, and to account for altruism towards deceased generations, appeared in Kimball (1987); Ray (1987); Hori and Kanaya (1989); Hori (1992); Fels and Zeckhauser (2008). Most closely related to our work however is the analysis in Bergstrom (1999) and Saez-Marti and Weibull (2005). Bergstrom considers general systems of interdependent preferences, demonstrating conditions under which these lead to well-behaved paternalistic preferences. He focuses largely on static, but heterogeneous, preferences, but does apply his framework to analyze the homogeneous two-sided first order PIA preferences in Kimball (1987). Saez-Marti and Weibull (2005), on the other hand, consider arbitrary forward-looking homogeneous intergenerational preferences. They identify conditions under which paternalistic time preferences can be mapped into PIA preferences in which each future generation is assigned positive weight. Our work synthesizes elements of these contributions – we allow both the heterogeneous preferences of Bergstrom (1999), and the forward-looking intergenerational preferences of Saez-Marti and Weibull (2005). The questions we ask are however different, as we will be chiefly concerned with the effects of heterogeneity on equivalent paternalistic time preferences, and how social preferences relate to individual preferences. Finally, Galperti and Strulovici (2015) have recently offered an axiomatization of forward-looking PIA. Their model is consistent with ours in the case where their intergenerational utility function (G in their notation) is positive affine, and preferences are homogeneous (see Section 3).

Although heterogeneous preferences have not been studied in the literature on pure intergenerational altruism, there is a sizeable literature on heterogeneous paternalistic time preferences. Gollier and Zeckhauser (2005) study utilitarian aggregation of additively separable time preferences, showing that in general representative time preferences will tend to the lowest discount rate in the population as the time horizon tends to infinity. This result has analogues in consumption savings models (Becker, 1980; Heal and Millner, 2013), and in models of market interactions when participants' time preferences are heterogenous (Lengwiler, 2005; Cvitanić et al., 2012). Jackson and Yariv (2014, 2015) study conceptual properties of time preference aggregation, showing that if agents' preferences are time invariant then the presence of heterogeneity forces social preferences to be time inconsistent.

Millner and Heal (2015) refine their findings, showing that if social preferences are utilitarian and individuals' preferences are discounted utilitarian, revealed preference cannot determine whether we should impose time consistency or time invariance on social preferences, but both are possible. They study the relationship between planning equilibria and decentralized allocation mechanisms in both cases, demonstrating that decentralization can strictly improve on social planning if preferences are time invariant. Although this literature deals with several related topics, it does not account for the possible dependence between agents' preferences, in contrast to the PIA literature discussed above. This paper is, so far as we are aware, the first to combine the conceptual elements of both these lines of work.

2 Model

The model is designed with normative public decision-making in mind, rather than the choices by private individuals. We assume N types of ethical agents, indexed by i, and that each type cares about the wellbeing of future societies in a unique manner. The distribution of types is assumed to be stable across generations. Our agents are ethical observers, who care about the whole of society, rather than a subset of future agents.[4] Let V_τ^i be the total wellbeing of type i in generation τ. In keeping with most of the literature, we assume that intergenerational preferences are additively separable (e.g. Saez-Marti and Weibull, 2005). Type i in generation τ assigns weight $f_s^i \geq 0$ to the wellbeing of generation $\tau + s$, where $s \geq 1$, and we assume that $\sum_{s=1}^{\infty} f_s^i < \infty$.[5] The societal consumption utility achieved by generation τ is U_τ, and is common to all types.

Let w_τ^i be the Pareto weight assigned to type i ($w_\tau^i > 0$, $\sum_j w_\tau^j = 1$) according to generation τ's social welfare function, which is assumed to be utilitarian, and thus additively separable in types. The level of social welfare attained by generation τ is thus given by

$$W_\tau := \sum_i w_\tau^i V_\tau^i. \tag{3}$$

[4]This is a sensitivity requirement on agents' preferences – each type should place positive weight on all types' preferences within a generation. This is a cross-sectional analogue of the requirement in homogeneous models of PIA that preferences be *direct*, i.e. they should place positive weight on each future generation (Galperti and Strulovici, 2015).

[5]This ensures that pure intergenerational preferences are summable when wellbeing levels are bounded. It does *not* automatically imply that wellbeing levels are bounded on bounded utility streams.

Given these assumptions, type i's intergenerational preferences are given by

$$\begin{aligned} V_\tau^i &= U_\tau + \sum_{s=1}^{\infty} f_s^i W_{\tau+s} \\ &= U_\tau + \sum_{s=1}^{\infty} f_s^i \left(\sum_j w_{\tau+s}^j V_{\tau+s}^j \right). \end{aligned} \tag{4}$$

In the special case where the weights f_s^i are common to all types this model reduces to the model of homogeneous forward-looking PIA in Saez-Marti and Weibull (2005), i.e.

$$V_\tau = U_\tau + \sum_{s=1}^{\infty} f_s V_{\tau+s}. \tag{5}$$

This homogeneous model is however paternalistic if agents' intergenerational preferences are heterogeneous. By contrast, the model in (4) is a fully non-paternalistic representation of intergenerational altruism, in that all types respect the preferences of other types when evaluating their wellbeing. Note that individuals with preferences (4) have not given up sovereignty over their own ethical views by being fully non-paternalistic. Type i's intergenerational altruism is captured by the set of weights f_s^i, which she is free to choose as she pleases. The difference between (4) and homogeneous models of PIA is that the *inputs* to the preference representation, i.e. the wellbeing of each type in future generations, are now computed according to their own preferences, and not according to the preferences of type i.[6]

In what follows we assume that each generation uses the same weights to aggregate types' wellbeing into a measure of social welfare at time τ:

$$\forall \tau, w_\tau^i = w^i. \tag{6}$$

This assumption requires a little explication. The Pareto weights w_τ^i that generation τ employs represent the outcome of its own ethical deliberations (whatever they may be) about how to aggregate the preferences of different contemporaneous types. Since the distribution of types is stable across generations, and since each type's preferences are forward looking, the system of preferences (4) is symmetric under translations of the time

[6] The preferences (4) can also be interpreted as a model of intra-personal time preferences, in which a single individual at time τ is uncertain about which preferences her 'future selves' at times $\tau + s$ will hold. In this case $w_{\tau+s}^i$ should be interpreted as self τ's subjective probability distribution over the preferences of self $\tau + s$.

axis $\tau \to \tau + \Delta$ for any $\Delta \in \mathbb{N}$. Thus if generation τ arrives at the Pareto weights w^i_τ as the solution to its preference aggregation problem, so will generation $\tau + s$ for any s. Thus (6) is a consequence of the symmetry of the preference system (4). In addition, non-paternalism rules out the possibility of the vector of Pareto weights being type dependent, since if it were, type i would be imposing it's own idea of how generation τ should aggregate preferences on τ, rather than respecting τ's own ethical deliberations on how to compute social wellbeing.

Given the interdependent preferences (4) our immediate task is to disentangle this recursive system to determine each type's equivalent paternalistic preferences over utility streams, as in the example in (1–2). In order to do so we need to solve this infinite system of coupled equations for the unknowns $V^i_{\tau+s}$ ($i \in \{1, \ldots, N\}, s \in \mathbb{N}$) in terms of the infinite utility sequence $(U_\tau, U_{\tau+1}, \ldots)$. Before we tackle this general problem it will be instructive to consider the simple example of first order PIA with heterogeneous preferences. This will allow us to anticipate several of the features of the more general case without the mathematical complications this entails.

2.1 Heterogeneous First Order PIA

If each type's intergenerational feelings extend only to their immediate descendants, the preference system (4) reduces to

$$V^i_\tau = U_\tau + \beta_i \left(\sum_j w^j V^j_{\tau+1} \right). \tag{7}$$

where we have defined $\beta_i := f^i_1$, and we assume $\beta_i \in [0, 1)$. Multiplying (7) through by w^i, and summing the N resulting equations, we find that social welfare W_τ satisfies the recurrence

$$W_\tau = U_\tau + \left(\sum_i w^i \beta_i \right) W_{\tau+1}. \tag{8}$$

Thus, social preferences W_τ exhibit homogeneous first order PIA. We know from (1–2) that this implies that

$$W_\tau = \sum_{s=0}^{\infty} \delta^s U_{\tau+s}, \text{ where } \delta := \sum_i w^i \beta_i. \tag{9}$$

Substituting this expression into (7), we find that type i's equivalent paternalistic preferences are given by

$$V^i_\tau = U_\tau + \beta_i \left(\sum_{s=1}^{\infty} \delta^{s-1} U_{\tau+s} \right). \tag{10}$$

These are nothing more than the quasi-hyperbolic preferences of e.g. Phelps and Pollak (1968); Laibson (1997), where the 'short run' discount factor β_i is unique to type i, but the 'long run' discount factor δ is common to all types, and equal to the weighted average of the β_i.

Several features of these simple results are worth highlighting. First, notice that if we set $\beta_i = \beta$ for all i, then $\delta = \beta$, and (10) reduces to the exponential paternalistic preferences of homogeneous first order PIA, as in (2). However, this is the only case in which the preferences (10) are exponential. This shows that the use of first order PIA to justify discounted utilitarian intergenerational time preferences (as in e.g. Barro (1974) and Barro and Becker (1989)), is non-robust to heterogeneity. Rather, quasi-hyperbolic individual time preferences are the generic consequence of first order PIA. Social preferences, however, are exponential. Thus, first order PIA gives rise to a population of present biased types, but a social welfare measure that is *not* present-biased. This time consistent welfare function is derived directly from agents' preferences, rather than being imposed in an *ad hoc* fashion.[7]

A second observation is that heterogeneous preferences have a 'mean reverting' effect on the paternalistic preferences of each type. By this we mean that each type's paternalistic time preferences are moderated towards the (weighted) mean, relative to the preferences they would have in a homogeneous model. We can see this directly in this model by writing out i's sequence of equivalent paternalistic discount factors:

$$\text{Homogeneous model: } 1, \beta_i, \beta_i^2, \beta_i^3, \ldots$$

$$\text{Heterogeneous model: } 1, \beta_i, \beta_i \left(\sum_j w^j \beta_j \right), \beta_i \left(\sum_j w^j \beta_j \right)^2, \ldots$$

Each element of the discount factor sequence for the heterogeneous model is weakly less

[7] Preferences are present biased if they exhibit more impatience towards payoffs that occur in the near future than they do for equivalent payoffs that occur in the far future. See e.g. Jackson and Yariv (2014) for a formal definition. It is well known that quasi-hyperbolic preferences exhibit present bias, whereas exponential preferences do not. There is a long tradition, beginning with Phelps and Pollak (1968), of imposing time consistent welfare measures on models of time inconsistent preferences. This has always been a somewhat dubious, and indeed paternalistic, move. Gul and Pesendorfer (2001) offer an alternative model of intertemporal choice that attempts to address this difficulty.

than the corresponding element for the homogeneous model if $\beta_i > \sum_j w^j \beta_j$, and vice versa if $\beta_i < \sum_j w^j \beta_j$. Thus the discount factor sequences of types with high discount factors are pulled downwards towards the weighted mean, and the discount factor sequences of types with low discount factors are pulled upwards towards the mean. Thus the homogeneous first order PIA model exhibits a systematic bias if preferences are heterogeneous and non-paternalistic.

Our final observation is that for $s \geq 2$, all types' weights on utility $U_{\tau+s}$ decline at the same geometric rate δ^s, where $\delta = \sum_j w^j \beta_j$. Moreover, this common discount factor corresponds to the *social* discount factor, i.e. the discount factor on utility streams according to social preferences $W_\tau = \sum_j w^j V_\tau^j$. Thus, although each type's preferences are idiosyncratic and unconstrained, all types agree on the appropriate discount *rate* to use in the long run.[8] A rough intuition for this finding follows from the fact that although each type's intergenerational weights are idiosyncratic, non-paternalism implies that all types respect the next generation's measure of social wellbeing, whatever it may be. All types' preferences are thus dependent on a common sequence of social preferences $W_{\tau+s}$. Types' preferences are additive in the social welfare of future generations, and since the distribution of preferences is invariant under time translations, social welfare decays at the same long-run rate in each future generation. Thus, each type's weights on future utilities must decay at the same long-run rate, since they are a linear combination of the social wellbeings of future generations, all of which decay at the same long-run rate. We will make this argument precise below.

While the first of these observations is specific to the first order model, which is of interest in its own right, both the second and third observations have analogues for higher order PIA models. We now turn to the general case.

2.2 The general case

We write type i's equivalent paternalistic time preferences as:

$$V_\tau^i = \sum_{s=0}^{\infty} a_s^i U_{\tau+s}. \tag{11}$$

[8] The long run discount rate ρ is defined through $\delta = (1+\rho)^{-1}$.

The coefficients a^i_s are unknowns, which must be determined by solving the system of preferences (4). In addition, we denote equivalent paternalistic social time preferences by

$$W_\tau = \sum_j w^j V^j_\tau = \sum_{s=0}^{\infty} b_s U_{\tau+s}, \text{ where } b_s = \sum_j w^j a^j_s. \tag{12}$$

The following proposition derives recursive formulas for a^i_s and b_s, which determine them as functions of the primitives w^i and f^i_s.

Proposition 1. *Define*

$$\lambda_s := \sum_j w^j f^j_s. \tag{13}$$

Then the social discount factor sequence b_s satisfies

$$b_s = \begin{cases} 1 & s = 0 \\ \sum_{r=1}^{s} \lambda_r b_{s-r} & s \geq 1. \end{cases} \tag{14}$$

Type i's discount factor sequence is related to the social discount factor through

$$a^i_s = \sum_{r=1}^{s} f^i_r b_{s-r}. \tag{15}$$

Proof. See Appendix A. □

Equation (15) shows that an efficient way to proceed in determining all types' equivalent paternalistic time preferences is to solve the single recurrence relation (14) for the social discount factor sequence, and then substitute the solution into (15) to determine each type's preferences.

2.2.1 Mean reversion of paternalistic preferences

Proposition 1 allows us to generalize our finding from the first order model that heterogeneity causes types' time preferences to 'mean revert'. To demonstrate this we need to put a little more structure on preferences:

Definition 1. The set of intergenerational weights $\{f^i_s\}$ is *well ordered* if:

$$\forall i, j > i \Rightarrow \forall s \in \mathbb{N}, f^i_s \geq f^j_s \tag{16}$$

where the inequality is strict for at least one value of s.

Intuitively, preferences that are well ordered can be arranged in increasing order of impatience.

Proposition 2. *Suppose that intergenerational preferences are well ordered. Define* $\tilde{V}^i_\tau$ *to be the 'homogeneous equivalent' preferences of type* i*, i.e.*

$$\tilde{V}^i_\tau = U_\tau + \sum_{s=1}^{\infty} f^i_s \tilde{V}^i_{\tau+s},$$

and let the equivalent paternalistic representation of preferences $\tilde{V}^i_\tau$ *be*

$$\tilde{V}^i_\tau = \sum_s \tilde{a}^i_s U_{\tau+s}.$$

Then,

$$f^i_1 \geq \sum_j w^j f^j_1 \Rightarrow \forall s \in \mathbb{N},\ b_s \leq a^i_s < \tilde{a}^i_s,$$
$$f^i_1 \leq \sum_j w^j f^j_1 \Rightarrow \forall s \in \mathbb{N},\ b_s \geq a^i_s > \tilde{a}^i_s.$$

Proof. See Appendix B. □

This result shows that the homogeneous model of PIA overestimates the weight an agent places on future utilities if they are more impatient than the (weighted) average, and underestimates it if they are less impatient than average.

2.2.2 Asymptotic agreement on discount rates

Our observation that all types agree on the 'long run' discount rate in the first order PIA model can be extended to all models of PIA of finite order, and to infinite order PIA models in which preferences are well behaved.

We begin by considering the finite order case. Assume that all types are altruistic towards at most the first M future generations, i.e.

$$M := \max\{s \in \mathbb{N} | \exists i \text{ s.t. } f^i_s > 0\} < \infty.$$

It follows that λ_s is non-zero only for $s \leq M$. The equation (14) for the social discount factor then becomes:

$$b_s = \sum_{r=1}^{M} \lambda_r b_{s-r}$$

We can solve this equation by making the guess $b_s = \mu^s$. Substituting into the expression, we find:

$$\mu^s = \sum_{r=1}^{M} \lambda_r \mu^{s-r}.$$

Multiply through by μ^{M-s} to find:

$$\mu^M - \sum_{r=1}^{M} \lambda_r \mu^{M-r} = 0. \tag{17}$$

This is an M-th order polynomial in μ, that will have M complex roots μ_k, $k = 1 \dots M$. By convention we will label the roots in non-increasing order of their complex modulus. A standard result in the theory of polynomials show that the largest root μ_1 is the unique positive root of (17) (see e.g. Lemma 8.1.1, p. 243, in Rahman and Schmeisser, 2002). Thus, $\mu_1 > |\mu_2| \geq \dots \geq |\mu_M|$. The general solution to (17) will be

$$b_s = \sum_{k=1}^{M} C_k (\mu_k)^s. \tag{18}$$

for some (possibly complex) constants C_k that are chosen to ensure that the first M elements of the sequence agree with the first $M - 1$ iterates of the recurrence, given the initial condition $b_0 = 1$.

Given the solution (18) for paternalistic social preferences, we can compute each type's paternalistic preferences from (15):

$$\begin{aligned} a_s^i &= \sum_{r=1}^{s} f_r^i \left(\sum_{k=1}^{M} C_k (\mu_k)^{s-r} \right) \\ &= \sum_{k} C_k (\mu_k)^s \sum_{r=1}^{s} f_r^i (\mu_k)^{-r} \end{aligned}$$

When $s > M$ this expression simplifies, since $f_r^i = 0$ for $r > M$. For $s > M$,

$$a_s^i = \sum_k C_k(\mu_k)^s \sum_{r=1}^{M} f_r^i(\mu_k)^{-r}. \tag{19}$$

We can write this more simply by defining

$$H_k^i(s) = C_k(\mu_k)^s \sum_{r=1}^{M} f_r^i(\mu_k)^{-r}. \tag{20}$$

Thus, for $s > M$,

$$a_s^i = \sum_{k=1}^{M} H_k^i(s). \tag{21}$$

The expression (19) shows that if $\mu_1 > 1$, there exist bounded utility streams on which V_τ^i is unbounded. This is clearly an undesirable property of preferences. For now we assume that the coefficients of (17) are such that $\mu_1 < 1$; this is the analogue of imposing $\beta < 1$ in the first order homogeneous model (2). We return later to deriving conditions on the primitives of the model that ensure that $\mu_1 < 1$.

With these calculations in place, we can prove the following:

Proposition 3. *If heterogeneous PIA preferences are of any finite order M, then for all i,*

$$\lim_{s\to\infty} \frac{a_{s+1}^i}{a_s^i} = \lim_{s\to\infty} \frac{b_{s+1}}{b_s} = \mu_1. \tag{22}$$

Proof. From (20–21), for $s > M$,

$$\begin{aligned} \frac{a_{s+1}^i}{a_s^i} &= \frac{\sum_{k=1}^{M} H_k^i(s+1)}{\sum_{k=1}^{M} H_k^i(s)} \\ &= \frac{\sum_{k=1}^{M} \mu_k H_k^i(s)}{\sum_{k=1}^{M} H_k^i(s)} \end{aligned} \tag{23}$$

Now

$$\begin{aligned} \lim_{s\to\infty} \frac{H_k^i(s)}{H_1^i(s)} &= \lim_{s\to\infty} \frac{C_k(\mu_k)^s \sum_{r=1}^{M} f_r^i(\mu_k)^{-r}}{C_1(\mu_1)^s \sum_{r=1}^{M} f_r^i(\mu_1)^{-r}} \\ &= \begin{cases} 0 & k \neq 1 \\ 1 & k = 1 \end{cases} \end{aligned}$$

since $\mu_1 > |\mu_k|$ for all $k \neq 1$. Dividing both the numerator and denominator of (23) by $H_1^i(s)$, and taking the limit as $s \to \infty$ we obtain $\lim_{s\to\infty} a_{s+1}^i/a_s^i = \mu_1$. Using the expression (18), we have

$$\frac{b_{s+1}}{b_s} = \frac{\sum_k C_k(\mu_k)^{s+1}}{\sum_k C_k(\mu_k)^s}.$$

Dividing both the numerator and denominator of this expression by $C_1(\mu_1)^s$, and taking the limit as $s \to \infty$, we obtain $\lim_{s\to\infty} b_{s+1}/b_s = \mu_1$. □

We can extend this result to the case of infinite order PIA provided that preferences are well behaved. Suppose that there is no $s \in \mathbb{N}$ such that for all $s' > s$, $f_{s'}^i = 0$ for all i. Thus the same property holds for λ_s. Define

$$\hat{V}_\tau^i = U_\tau + \sum_{s=1}^{\infty} f_s^i(\sum_j w^j \hat{V}_{\tau+s}^j)$$

and these are assumed to have the paternalistic representation

$$\hat{V}_\tau^i = \sum_{s=0}^{\infty} \hat{a}_s^i U_{\tau+s}. \tag{24}$$

As before, define $\hat{b}_s = \sum_j w^j \hat{a}_s^j$.

Proposition 4. *Let $\mu_1(M)$ to be the unique positive root of the equation*

$$x^M - \sum_{r=1}^{M} \lambda_r x^{M-r} = 0, \tag{25}$$

Then the sequence $\mu_1(M)$ is increasing, and $\hat{\mu}_1 := \lim_{M\to\infty} \mu_1(M)$ exists. If $\hat{\mu}_1 < 1$,

$$\lim_{s\to\infty} \frac{\hat{a}_{s+1}^i}{\hat{a}_s^i} = \lim_{s\to\infty} \frac{\hat{b}_{s+1}}{\hat{b}_s} = \hat{\mu}_1. \tag{26}$$

Proof. See Appendix C. □

These results have something of the flavour of related findings on the aggregation of uncertain real interest rates (Weitzman, 1998, 2001; Freeman and Groom, 2014), and on the aggregation of pure time preferences in standard paternalistic models (e.g. Gollier and Zeckhauser, 2005; Millner and Heal, 2015). In each of these cases averaging over a distribution of discount factors leads to a 'certainty equivalent' discount rate, or a representative discount rate, that declines to the lowest rate as the time horizon tends to infinity. We also find that the long-run discount factor on utility streams is the largest exponential component of (18), i.e. the component that declines at the lowest rate. Unlike these other findings however, which all pertain to an asymptotic property of some aggregator (either representative preferences or certainty equivalents), our finding applies to individuals' preferences themselves. This is a much stronger convergence property, which is intimately related to the interdependence between the preferences of different types.

2.2.3 Well behaved paternalistic preferences

The condition $\hat{\mu}_1 < 1$ in Proposition 4 ensures that infinite order preferences $\hat{V}_\tau^i$ are bounded when the utility stream $(U_\tau, U_{\tau+1}, \ldots)$ is bounded. This is clearly a desirable property, in both finite and infinite order models. Yet this condition is not guaranteed to hold for all preferences – we can easily find coefficients of the polynomial (17) such that $\mu_1 > 1$. In addition to the possibility of preferences being unbounded, there is nothing in the solution (18) that prevents a_s^i from being non-positive for some s. In this case a types' wellbeing would *increase* if $U_{\tau+s}$ were diminished. This is also an undesirable feature. Finally, we also do not know whether equivalent paternalistic time preferences decline monotonically to 0, as we would expect from well-behaved preferences. It is thus of interest to identify a set of conditions on the primitives of the model that ensures that paternalistic preferences are well behaved. The following proposition provides these:

Proposition 5. *(a) Suppose that:*

$$\forall i, \forall M \geq 1 \sum_{s=0}^{M} f_s^i < 1. \tag{27}$$

Then for all i,

i) V_τ^i is increasing in $U_{\tau+s}$ for all s, i.e. $a_s^i \geq 0$ for all $s \in \mathbb{N}$.

ii) $\hat{\mu}_1 < 1$. Thus all types agree on the asymptotic discount rate, and $V_\tau^i < \infty$ for all bounded utility streams, i.e. $\lim_{s\to\infty} a_s^i = 0$.

(b) If $\{f_r^i\}$ is well ordered, $f_1^i < 1$ for all i, and

$$\forall i, \forall r \geq 1, \; f_{r+1}^i < (1 - f_1^i) f_r^i, \tag{28}$$

then (27) is satisfied, so $b_s > 0, a_s^i > 0, \lim_{s\to 0} a_s^i = 0 = \lim_{s\to 0} b_s$*. In addition,* b_s *is monotonically decreasing in* s*, and* a_s^i *is monotonically decreasing in* s *if* $f_1^i > \lambda_1$*.*

Proof. See Appendix D. □

The results of this proposition hold for models of arbitrary order, finite or infinite. Although it is of interest to know whether all types' paternalistic preferences a_s^i are decreasing in s, the partial result in part (b) of the proposition is the strongest we've obtained. If we are interested only in whether social preferences b_s are monotonically declining, the proof of part (b) of the proposition provides a sufficient condition to ensure this:

$$\lambda_1 < 1, \lambda_{r+1} \leq (1-\lambda_1)\lambda_r \text{ for all } r \Rightarrow b_s \geq b_{s+1} \text{ for all } s. \tag{29}$$

For this inequality to be satisfied, it is necessary for

$$\lambda_r \leq (1-\lambda_1)^{r-1}\lambda_1. \tag{30}$$

Thus λ_r must decline at least exponentially, with a minimum rate $(1-\lambda_1)^r$. The larger is λ_1, the faster λ_r must fall in order for (29) to be satisfied. Similarly, for (28) to be satisfied f_r^i must decline at a minimum rate $(1-f_1^i)^r$.

3 Quasi-hyperbolic preferences and heterogeneous PIA

In this section we apply our model to the axiomatic characterization of intergenerational altruism in Galperti and Strulovici (2015). By adapting the axiomatic treatment of paternalistic time preferences in Koopmans (1960) to a model of PIA, they derive the following representation for preferences:

$$V_\tau = U_\tau + \sum_{s=1}^{\infty} \alpha^s G(V_{\tau+s}). \tag{31}$$

where $\alpha \in [0,1)$ and G is an increasing function. If they add an additional axiom, which they call Consumption Independence[9], the function G must be linear, i.e.

$$V_\tau = U_\tau + \sum_{s=1}^{\infty} \alpha^s \gamma V_{\tau+s}. \tag{32}$$

[9] This condition essentially requires the marginal rate of substitution between consumption at times $\tau+s, \tau+s'$ to be independent of consumption at any other time $\tau+s''$, where $s'' \notin \{s, s'\}$.

where $\gamma > 0$. Using (14) it is straightforward to show that these preferences have the paternalistic representation

$$V_\tau = U_\tau + \beta \sum_{s=1}^{\infty} \delta^s U_{\tau+s}, \tag{33}$$

$$\beta = \frac{\gamma}{\gamma+1};\ \delta = (1+\gamma)\alpha. \tag{34}$$

Thus, the paternalistic preferences corresponding to their axiomatically motivated model of PIA are quasi-hyperbolic. Galperti and Strulovici (2015) thus view their model as providing a normative justification for the quasi-hyperbolic discounting model of Phelps and Pollak (1968) in intergenerational decision problems.

Under the axioms of their model the form of individuals' PIA preferences are prescribed, yet these individuals are paternalistic if preferences are heterogeneous. We thus consider a model in which each type i's preferences obey their axioms, but operate on the social wellbeing $W_{\tau+s}$ of future generations, rather than on the wellbeing of only type i:

$$V_\tau^i = U_\tau + \gamma_i \sum_{s=1}^{M} \alpha_i^s \left(\sum_j w^j V_{\tau+s}^j \right), \tag{35}$$

where α_i, γ_i are idiosyncratic to type i, and we allow the order of preferences M to be finite or infinite. As $M \to \infty$ we recover the preferences of Galperti and Strulovici (2015). In our notation, this model corresponds to the choice

$$f_s^i = \begin{cases} \gamma_i \alpha_i^s & s \leq M \\ 0 & s > M. \end{cases} \tag{36}$$

$$\Rightarrow \lambda_s = \sum_j w^j \gamma_j \alpha_j^s. \tag{37}$$

From our results (18) and (19), it is clear that in general individual types' equivalent paternalistic preferences will not be quasi-hyperbolic in this model. If preferences are heterogeneous the only case in which types' preferences are quasi-hyperbolic is the case $M = 1$, and not the $M = \infty$ case prescribed by the axioms of Galperti and Strulovici (2015). Moreover, in this case, social preferences are exponential, not quasi-hyperbolic. Can we construct a model in which social preferences are quasi-hyperbolic, even if types' preferences are not? From (32–33), we see that this would require the sequence λ_s to obey

$$\lambda_s = \gamma\alpha^s \tag{38}$$

for some γ and α. But since $\lambda_s = \sum_i w^i \gamma_i \alpha_i^s$, this can only be true in a homogeneous model. Thus, quasi-hyperbolic social preferences are generically ruled out if there is any heterogeneity in preferences.

We can use Proposition 5 to get a handle on the behavior of equivalent paternalistic preferences in this model. We have

$$\begin{aligned}\sum_{s=1}^{M} \lambda_s &= \sum_{s=1}^{M} \sum_j w^j \gamma_j \alpha_j^s \\ &= \sum_j w^j \gamma_j \frac{\alpha_j(1-\alpha_j^M)}{1-\alpha_j} \end{aligned}\tag{39}$$

Thus we can apply part (a) of Proposition 5 for all M if

$$\sum_i w^j \gamma_j \frac{\alpha_j}{1-\alpha_j} < 1. \tag{40}$$

Similarly, part (b) of Proposition 5 holds if

$$\begin{aligned} & f^i_{r+1} < (1-f^i_1) f^i_r \\ & \iff \gamma_i \alpha_i^{r+1} < (1-\gamma_i\alpha_i)\gamma_i\alpha_i^r \\ & \iff \gamma_i < \frac{1-\alpha_i}{\alpha_i}. \end{aligned}\tag{41}$$

So if (41) holds so does (40), and all the conclusions of Proposition 5 follow. For example, if $\gamma^j = 1$ for all j, the conclusions of Proposition 5 hold provided that $\alpha_j < 1/2$ for all j. In particular, all types will agree on the long-run discount rate despite the differences in their intergenerational preferences in both finite and infinite order versions of the model. Figure 1 demonstrates the behaviour of types' paternalistic discount factors in this model, showing that paternalistic preferences are not quasi-hyperbolic for any $M > 1$.

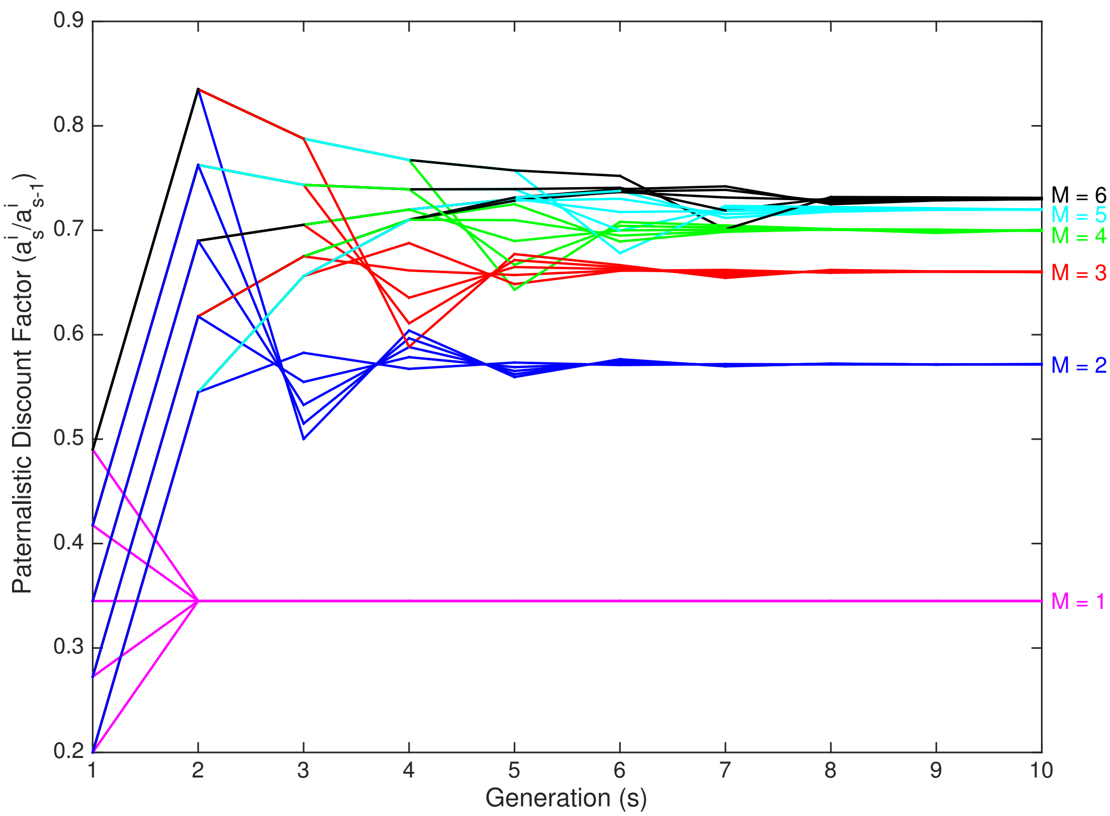

Figure 1: Equivalent paternalistic discount factors a^i_s/a^i_{s-1} for models of order $M = 1 \dots 6$, with $N = 5$ types. $\gamma^j = 1$ for all j, the five values α_i are chosen to be uniform on $[0.2, 0.49]$, and $w^i = 1/5$ for all i. The figure shows the convergence of types' discount factors in the long run, and the fact that types' paternalistic preferences are not quasi-hyperbolic for any $M > 1$. Since all discount factors are less than 1, a^i_s is decreasing in s for all s in this model. The limiting value of the long run discount factor in the infinite order model, $\hat{\mu}_1 = \lim_{M\to\infty} \mu_1(M)$, is already well approximated by $\mu_1(6)$.

4 Conclusions

This paper introduced preference heterogeneity into the literature on pure intergenerational altruism. Our core finding is that heterogeneity qualitatively alters the findings of existing homogeneous models of PIA. In particular, derivations of exponential or quasi-hyperbolic time preferences from these models do not hold up if preferences are heterogeneous. In general, homogeneous models under (over) estimate the weight a non-paternalistic agent places on future utilities if she is more (less) impatient than a socially weighted average. Our strongest result shows that if preferences are heterogeneous, and agents are fully non-paternalistic towards future societies, then all agents agree on the asymptotic utility discount rate, regardless of their preferences. Thus in the long run the entanglement between non-paternalistic agents' preferences washes away all heterogeneity in discount rates.

The fact that agents with diverse ethical views agree on long-run discount rates in our model is particularly striking, given persistent disagreements about appropriate values for the ethical parameters of social discount rates in the literature (Drupp et al., 2014). These disagreements are especially consequential for economic analyses of long run policy questions such as climate change (e.g. Stern, 2007; Nordhaus, 2008). Since most of the benefits of current greenhouse gas abatement policies will only be realized by future generations, the choice of long-run discount rates has an enormous effect on cost-benefit analyses of climate policy. A 2% disagreement about the appropriate value of the utility discount rate, for example, can cause recommendations of first-best carbon taxes to vary by an order of magnitude. Our model suggests one possible route out of this impasse. If we believe that current ethical disagreements about societal discount rates are likely to be replicated in the future, and agents with heterogeneous views can agree to act non-paternalistically to achieve consensus (while reserving the right to recommend any normative discounting schedule they please), the model yields a unique long-run utility discount rate, which all agents must agree on. If an agent wishes to deviate from this consensus value after it has been produced she would be guilty of paternalism, a stance that must surely affect her standing in any consensus building exercise. Taking non-paternalism to its logical conclusion may thus be a powerful consensus builder among economists with heterogeneous views on social discounting. Such a consensus on key normative parameters could provide a welcome shot in the arm for the legitimacy of cost-benefit analyses of long-run policies.

References

Arrow, K. J. (1999). Discounting, morality, and gaming. In P. R. Portney and J. P. Weyant (Eds.), *Discounting and Intergenerational Equity*. Resources for the Future.

Barro, R. J. (1974). Are Government Bonds Net Wealth? *Journal of Political Economy 82*(6), 1095–1117.

Barro, R. J. and G. S. Becker (1989). Fertility Choice in a Model of Economic Growth. *Econometrica 57*(2), 481–501.

Becker, G. S. (1974). A Theory of Social Interactions. *Journal of Political Economy 82*(6), 1063–1093.

Becker, R. A. (1980). On the Long-Run Steady State in a Simple Dynamic Model of Equilibrium with Heterogeneous Households. *The Quarterly Journal of Economics 95*(2), 375–382.

Bergstrom, T. C. (1999). Systems of Benevolent Utility Functions. *Journal of Public Economic Theory 1*(1), 71–100.

Cvitanić, J., E. Jouini, S. Malamud, and C. Napp (2012). Financial Markets Equilibrium with Heterogeneous Agents. *Review of Finance 16*(1), 285–321.

Dasgupta, P. S. and G. M. Heal (1979). *Economic Theory and Exhaustible Resources.* Cambridge University Press.

Drupp, M., M. Freeman, B. Groom, and F. Nesje (2014). Discounting disentangled: An expert survey on the determinants of the long-term social discount rate. *Mimeo, London School of Economics and Political Science*.

Fels, S. and R. Zeckhauser (2008). Perfect and total altruism across the generations. *Journal of Risk and Uncertainty 37*(2-3), 187–197.

Fishburn, P. C. and A. Rubinstein (1982). Time Preference. *International Economic Review 23*(3), 677–694.

Fisher, I. (1926). *Mathematical investigations in the theory of value and price.* Yale University Press.

Freeman, M. C. and B. Groom (2014). Positively gamma discounting: combining the opinions of experts on the social discount rate. *The Economic Journal (forthcoming).*

Galperti, S. and B. Strulovici (2015). A Theory of Intergenerational Altruism. *Working Paper:* ***http://faculty.wcas.northwestern.edu/~bhs675/TIC.pdf***.

Gollier, C. and R. Zeckhauser (2005). Aggregation of Heterogeneous Time Preferences. *Journal of Political Economy 113*(4), 878–896.

Gul, F. and W. Pesendorfer (2001). Temptation and Self-Control. *Econometrica 69*(6), 1403–1435.

Heal, G. and A. Millner (2013). Discounting under disagreement. *NBER Working paper no. 18999*.

Hori, H. (1992). Utility functionals with nonpaternalistic intergenerational altruism: The case where altruism extends to many generations. *Journal of Economic Theory 56*(2), 451–467.

Hori, H. and S. Kanaya (1989). Utility functionals with nonpaternalistic intergenerational altruism. *Journal of Economic Theory 49*(2), 241–265.

Jackson, M. O. and L. Yariv (2014). Present Bias and Collective Dynamic Choice in the Lab. *American Economic Review 104*(12), 4184–4204.

Jackson, M. O. and L. Yariv (2015). Collective Dynamic Choice: The Necessity of Time Inconsistency. *American Economic Journal: Microeconomics 7*(4), 150–78.

Kimball, M. S. (1987). Making sense of two-sided altruism. *Journal of Monetary Economics 20*(2), 301–326.

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

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics 112*(2), 443–478.

Lengwiler, Y. (2005). Heterogeneous Patience and the Term Structure of Real Interest Rates. *The American Economic Review 95*(3), 890–896.

Millner, A. and G. Heal (2015). Collective Intertemporal Choice: Time Consistency vs. Time Invariance. *Working Paper*.

Nordhaus, W. D. (2008). *A question of balance*. Yale University Press.

Pearce, D. G. (2008, November). Nonpaternalistic sympathy and the inefficiency of consistent intertemporal plans. In M. O. Jackson and A. McLennan (Eds.), *Foundations in Microeconomic Theory: A Volume in Honor of Hugo F. Sonnenschein*. Springer Science & Business Media.

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

Pigou, A. C. (1903). Some Remarks on Utility. *The Economic Journal 13*(49), 58–68.

Rahman, Q. I. and G. Schmeisser (2002). *Analytic Theory of Polynomials*. Clarendon Press.

Ray, D. (1987). Nonpaternalistic intergenerational altruism. *Journal of Economic Theory 41*(1), 112–132.

Saez-Marti, M. and J. W. Weibull (2005). Discounting and altruism to future decision-makers. *Journal of Economic Theory 122*(2), 254–266.

Sen, A. (2010). *The idea of justice*. London: Penguin.

Stern, N. H. (2007). *The economics of climate change : the Stern review*. Cambridge: Cambridge University Press.

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies 23*(3), 165.

Weitzman, M. L. (1998). Why the Far-Distant future should be discounted at its lowest possible rate,. *Journal of Environmental Economics and Management 36*(3), 201 – 208.

Weitzman, M. L. (2001). Gamma discounting. *The American Economic Review 91*(1), 260–271.

Weitzman, M. L. (2007). A Review of The Stern Review on the Economics of Climate Change. *Journal of Economic Literature 45*(3), 703–724.

Yaari, M. E. (1965). Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *The Review of Economic Studies 32*(2), 137–150.

A Proof of Proposition 1

Substitute (11) into (4):

$$\sum_{s=0}^{\infty} a_s^i U_{\tau+s} = U_\tau + \sum_{r=1}^{\infty} f_r^i \left(\sum_j w_j \sum_{q=0}^{\infty} a_q^j U_{\tau+r+q} \right)$$

Equating the coefficients of $U_{\tau+s}$ on the left and right hand sides of this expression we find

$$a_0^i = 1 \tag{42}$$

$$\begin{aligned} a_s^i &= \sum_{r\geq 1, q\geq 0, r+q=s} f_r^i \left(\sum_j w_j a_q^j \right) \\ &= \sum_{r=1}^{s} f_r^i \left(\sum_j w_j a_{s-r}^j \right) \\ &= \sum_{r=1}^{s} f_r^i b_{s-r} \end{aligned} \tag{43}$$

Now multiply (43) by w_i:

$$w_i a_s^i = \sum_{r=1}^{s} w_i f_r^i b_{s-r}$$

Thus

$$b_s = \sum_i w_i a_s^i = \sum_i \sum_{r=1}^{s} w_i f_r^i b_{s-r} = \sum_{r=1}^{s} \lambda_r b_{s-r}.$$

B Proof of Proposition 2

Using manipulations identical to those in the proof of Proposition 1, one can show that the homogeneous equivalent paternalistic preferences satisfy

$$\tilde{a}_s^i = \sum_{r=1}^{s} f_r^i \tilde{a}_{s-r}, \tag{44}$$
$$\tilde{a}_0^i = 1.$$

From Proposition (1), we know that

$$a_s^i = \sum_{r=1}^{s} f_r^i b_{s-r} \tag{45}$$
$$b_s = \sum_{r=1}^{s} \lambda_r b_{s-r} \tag{46}$$
$$a_0^i = b_0 = 1.$$

Compare the two recurrence relations (44) and (46). Since $a_0^i = b_0 = 1$, it is clear that if $f_r^i \geq \lambda_r$ for all r, we must have $\tilde{a}_s \geq b_s$ for all s. Now comparing (44) and (45), if $\tilde{a}_s \geq b_s$ for all s, it follows that $a_s^i < \tilde{a}_s^i$. Thus, $f_r^i \geq \lambda_r$ for all r implies $a_s^i < \tilde{a}_s^i$. Since preferences are well ordered, $f_1^i \geq \lambda_1 \Rightarrow f_r^i \geq \lambda_r$ for all $r > 1$, since $\lambda_r = \sum_j w^j f_r^j$. Finally, inspection of (45–46) shows that $b_s \leq a_s^i$ when $f_r^i \geq \lambda_r$ for all r . Analogous reasoning holds when $f_1^i \leq \lambda_1$.

C Proof of Proposition 4

Define

$$V_\tau^i(M) = U_\tau + \sum_{s=1}^{M} f_s^i (\sum_j w^j V_{\tau+s}^j(M)) \tag{47}$$

which has an associated paternalistic representation

$$V_\tau^i(M) = \sum_{s=0}^{\infty} a_s^i(M) U_{\tau+s}, \tag{48}$$

and we also define $b_s(M) = \sum_j w^j a_s^j(M)$. By the definition of an infinite sum, infinite order preferences are the formal limit of finite order preferences as $M \to \infty$:

$$\hat{V}_\tau^i = \lim_{M\to\infty} \left[U_\tau + \sum_{s=1}^{M} f_s^i(\sum_j w^j V_{\tau+s}^j(M)) \right], \tag{49}$$

We begin by proving the following lemma:

Lemma 1.

$$\hat{\mu}_1 := \lim_{M\to\infty} \mu_1(M) \ \textit{exists} \ .$$

Proof. Write the equation (25) as $L_M(x) = R_M(x)$, where $L_M(x)$ and $R_M(x)$ are its left and right hand sides respectively. Both $L_M(x)$ and $R_M(x)$ are increasing and positive for $x \geq 0$, and $L_M(0) = 0 < R_M(0) = \lambda_M$. Thus $L_M(x) < R_M(x)$ for $x \in [0, \mu_1(M))$, and $L_M(x) > R_M(x)$ for $x > \mu_1(M)$. Now consider the corresponding equation for $\mu_1(M+1)$:

$$x^{M+1} = \sum_{r=1}^{M+1} \lambda_r x^{M+1-r}.$$

For $x \neq 0$, we can divide this equation by x to find that it is equivalent to

$$L_M(x) = R_M(x) + \lambda_{M+1}/x. \tag{50}$$

The right hand side of (50) is strictly greater than $R_M(x)$ for positive x, which is itself greater than $L_M(x)$ for $x < \mu(M)$. Thus the positive solution of (50) must exceed $\mu_1(M)$, i.e. $\mu_1(M+1) > \mu_1(M)$. A corollary of a theorem of Cauchy (see Theorem 8.1.3, and Corollary 8.1.8, in Rahman and Schmeisser (2002)) shows that

$$\mu_1(M) < 1 + \max_{s\leq M} \lambda_s \tag{51}$$

Since λ_s is bounded for all $s \in \mathbb{N}$, the sequence $\mu_1(M)$ is bounded above. Thus since $\mu_1(M)$ is increasing and bounded above, the monotone convergence theorem guarantees that $\lim_{M\to\infty} \mu_1(M)$ exists. □

We proved in Proposition 3 that

$$\lim_{M\to\infty} \lim_{s\to\infty} \frac{a_{s+1}^i(M)}{a_s^i(M)} = \lim_{M\to\infty} \lim_{s\to\infty} \frac{b_{s+1}(M)}{b_s(M)} = \hat{\mu}_1. \tag{52}$$

We now wish to know whether under the conditions of this proposition it is also true that:

$$\lim_{s\to\infty}\lim_{M\to\infty}\frac{a^i_{s+1}(M)}{a^i_s(M)} = \lim_{s\to\infty}\lim_{M\to\infty}\frac{b_{s+1}(M)}{b_s(M)} = \hat{\mu}_1. \tag{53}$$

That is, can we change the order of the limits in (52)? In general, for limit operations to be interchangeable we require the sequence of functions they operate on to be *uniformly convergent*. The functions in question here are $V^i_\tau(M)$ and $\hat{V}^i_\tau$, which we can think of as linear functions from the infinite dimensional space $\mathbb{R}^\infty = \{\vec{U} = (U_\tau, U_{\tau+1}, U_{\tau+2}, \ldots)|\forall s \in \mathbb{N}, U_{\tau+s} \in \mathbb{R}\}$ to $\mathbb{R}$. If the sequence of functions $V^i_\tau(M)$ converges uniformly to $\hat{V}^i_\tau$ on any bounded subset of $\mathbb{R}^\infty$, then (53) will be satisfied. We now prove a second lemma:

Lemma 2. *Let B be a compact subset of $\mathbb{R}^\infty$, and assume that $\hat{\mu}_1 < 1$. Then $V^i_\tau(M)$ converges uniformly to $\hat{V}^i_\tau$ on B.*

Proof. Equation (43) shows that for all $s \le M$, $a^i_{\tau+s}(M) = \hat{a}^i_{\tau+s}$. Let $\bar{U} = \sup_s\{\sup\{U_{\tau+s}|U_{\tau+s} \in B\}\}$. For any $\vec{U} \in B$,

$$\begin{aligned}
\sup_{\vec{U}\in B}\left|V^i_\tau(M) - \hat{V}^i_\tau\right| &= \sup_{\vec{U}\in B}\left|\sum_{s=1}^{\infty} a^i_{\tau+M+s}(M)U_{\tau+M+s} - \sum_{s=1}^{\infty}\hat{a}^i_{\tau+M+s}U_{\tau+M+s}\right| && (54)\\
&\le \sum_{s=1}^{\infty}\left[\left|a^i_{\tau+M+s}(M)\right| + \left|\hat{a}^i_{\tau+M+s}\right|\right]\bar{U} && (55)
\end{aligned}$$

By Lemma 1, $\hat{\mu}_1 < 1$ also implies $\mu_1(M) < 1$ for all M, so we know that $\lim_{M\to\infty} a^i_{\tau+M+s}(M) = 0 = \lim_{M\to\infty}\hat{a}^i_{\tau+M+s}$. Thus

$$\lim_{M\to\infty}\sup_{\vec{U}\in B}\left|V^i_\tau(M) - \hat{V}^i_\tau\right| = 0.$$

Hence $V^i_\tau(M)$ converges uniformly to $\hat{V}^i_\tau(M)$. □

This completes the proof.

D Proof of Proposition 5

Proof of part (a):

i) Define:

$$\vec{V}_\tau^i = \begin{pmatrix} V_\tau^i \\ V_{\tau+1}^i \\ V_{\tau+2}^i \\ \vdots \end{pmatrix}, \quad \vec{W}_\tau = \sum_i w^i \vec{V}_\tau^i = \begin{pmatrix} W_\tau \\ W_{\tau+1} \\ W_{\tau+2} \\ \vdots \end{pmatrix}, \quad \vec{U}_\tau = \begin{pmatrix} U_\tau \\ U_{\tau+1} \\ U_{\tau+2} \\ \vdots \end{pmatrix}.$$

$$\mathbf{F}^i = \begin{pmatrix} 0 & f_1^i & f_2^i & f_3^i & \dots \\ 0 & 0 & f_1^i & f_2^i & \dots \\ 0 & 0 & 0 & f_1^i & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \mathbf{\Lambda} = \sum_i w^i \mathbf{F}^i = \begin{pmatrix} 0 & \lambda_1 & \lambda_2 & \lambda_3 & \dots \\ 0 & 0 & \lambda_1 & \lambda_2 & \dots \\ 0 & 0 & 0 & \lambda_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Using (15), we can rewrite the system of preferences (4) as the following system of coupled matrix equations:

$$\vec{W}_\tau = \vec{U}_\tau + \mathbf{\Lambda}\vec{W}_\tau, \tag{56}$$

$$\vec{V}_\tau^i = \vec{U}_\tau + \mathbf{F}^i\vec{W}_\tau. \tag{57}$$

The solution to (56) is

$$\vec{W}_\tau = (\mathbf{1} - \mathbf{\Lambda})^{-1}\vec{U}_\tau, \tag{58}$$

where $\mathbf{1}$ is the infinite dimensional identity matrix. Proposition 2 in Bergstrom (1999) shows that the solution $\vec{W}_\tau$ exists, is unique, and its elements are increasing in $U_{\tau+s}$ for all $s \in \mathbb{N}$, if $\mathbf{1}-\mathbf{\Lambda}$ is a dominant diagonal matrix. A matrix $\mathbf{A}$ is dominant diagonal if its elements A_{ij} satisfy $|A_{ii}| > \sum_{j\neq i} |A_{ij}|$ for all i. Thus, $\mathbf{1}-\mathbf{\Lambda}$ is dominant diagonal if $\sum_s \lambda_s < 1$. Since $\lambda_s = \sum_i w^i f_s^i$, this condition is satisfied if $\sum_s f_s^i < 1$ for all i. Now since $\mathbf{F}^i > 0$, (57) implies that if W_τ is increasing in $U_{\tau+s}$ for all s, so is V_τ^i for all i.

ii) Since $\sum_{s=1}^\infty f_s^i < 1$ for all i, $\sum_s \lambda_s < 1$. For finite M, it is well known that the moduli of the roots of the equation (17) are strictly less than $\max\{1, \sum_s \lambda_s\}$, which is equal to 1 by assumption. Thus $\mu_1(M) < 1$ for all finite M. We prove that $\hat{\mu} = \lim_{M\to\infty} \mu(M) < 1$. Dividing (17) through by μ^M, we see that $\hat{\mu}$ must be a solution of

$$1 = \sum_{r=1}^{\infty} \lambda_r \mu^{-r}. \tag{59}$$

This equation cannot be satisfied at $\mu = 1$, since $\sum_{s=1}^\infty \lambda_s < 1$. Thus by Proposition 4, all types agree on asymptotic discount rates. In addition, since a_s^i declines geometrically as $s \to \infty$, $\lim_s a_s^i = 0$.

Proof of part (b):

We begin by showing that if (28) holds, and $\{f_r^i\}$ are well ordered, then

$$\lambda_{r+1} < (1 - \lambda_1)\lambda_r. \tag{60}$$

Multiplying (28) through by w_i, and summing the resulting N inequalities, we find

$$\begin{aligned} \lambda_{r+1} &< \lambda_r - \sum_i w^i f_1^i f_r^i \\ &= \lambda_r(1 - \lambda_r) + \lambda_r \lambda_1 - \sum_i w^i f_1^i f_r^i. \end{aligned}$$

Now

$$\begin{aligned} \lambda_r \lambda_1 - \sum_i w^i f_1^i f_r^i &= (\sum_j w^j f_r^j)(\sum_k w^k f_1^k) - \sum_i w^i f_1^i f_r^i \\ &= (\mathsf{E} f_r^j)(\mathsf{E} f_r^j) - (\mathsf{E} f_r^j) \\ &= -\mathrm{Cov}_i(f_1^i, f_r^i) \end{aligned}$$

where the operator E denotes an expectation over types i with probability distribution w^i, and Cov is the covariance. Since f_r^i is well ordered, $\mathrm{Cov}_i(f_1^i, f_r^i) > 0$, and thus (60) holds.

When (60) holds, it follows that

$$\lambda_r \leq (1 - \lambda_1)^{r-1} \lambda_1. \tag{61}$$

Thus

$$\begin{aligned} \sum_{r=1}^{s} \lambda_s &\leq \sum_{r=1}^{s} \lambda_1 (1 - \lambda_1)^{r-1} \\ &= \lambda_1 \frac{1 - (1 - \lambda_1)^s}{1 - (1 - \lambda_1)} \\ &= 1 - (1 - \lambda_1)^s. \end{aligned}$$

By assumption, $\lambda_1 < 1$. Hence

$$\sum_{r=1}^{s} \lambda_s < 1.$$

From part (a) of this proposition, we conclude that $a_i^s, b_s > 0$ for all s, and $\lim_{s\to\infty} a_s^i = \lim_{s\to\infty} b_s = 0$.

We now show that (60) implies that b_s is decreasing. From (14),

$$\begin{aligned} b_s &= \sum_{r=1}^{s} \lambda_r b_{s-r} \\ b_{s+1} &= \sum_{r=1}^{s+1} \lambda_r b_{s+1-r} \\ &= \lambda_1 b_s + \sum_{r=1}^{s} \lambda_{r+1} b_{s-r}. \end{aligned}$$

So,

$$\begin{aligned} b_s - b_{s+1} &= -\lambda_1 b_s + \sum_{r=1}^{s} (\lambda_r - \lambda_{r+1}) b_{s-r} \\ &= -\lambda_1 \sum_{r=1}^{s} \lambda_r b_{s-r} + \sum_{r=1}^{s} (\lambda_r - \lambda_{r+1}) b_{s-r} \\ &= \sum_{r=1}^{s} [(1-\lambda_1)\lambda_r - \lambda_{r+1}] b_{s-r}. \end{aligned}$$

Since $b_{s-r} > 0$ and $[(1-\lambda_1)\lambda_r - \lambda_{r+1}] > 0$ when (60) is satisfied, $b_s > b_{s+1}$.

Similarly, since

$$a_s^i = \sum_{r=1}^{s} f_r^i b_{s-r}$$

it follows that

$$\begin{aligned} a_s^i - a_{s+1}^i &= -f_1^i b_s + \sum_{r=1}^{s} (f_r^i - f_{r+1}^i) b_{s-r} \\ &= -f_1^i \sum_{r=1}^{s} \lambda_r b_{s-r} + \sum_{r=1}^{s} (f_r^i - f_{r+1}^i) b_{s-r} \\ &= \sum_{r=1}^{s} [f_r^i - f_{r+1}^i - f_1^i \lambda_r] b_{s-r}. \end{aligned}$$

Since preferences are well ordered, $f_1^i > \lambda_1 \Rightarrow f_r^i > \lambda_r$. Thus for i such that $f_1^i > \lambda_1$,

$$a_s^i - a_{s+1}^i > \sum_{r=1}^{s} [f_r^i - f_{r+1}^i - f_1^i f_r^i] b_{s-r} > 0 \tag{62}$$

where the last inequality follows from the fact that $b_s > 0$ and our assumption (28). Thus a_s^i decreases monotonically if $f_1^i > \lambda_1$.