Systems of Benevolent Utility Functions

THEODORE C. BERGSTROM
University of California at Santa Barbara

Abstract

This paper studies systems of utility functions in which each person's utility depends on his or her own consumption as well as on the utilities of others. We consider the question of when a system of interdependent utility functions induces unique utility functions over allocations and identifies the class of transformations on interdependent utility functions that are equivalent in the sense of inducing the same preferences over allocations. We show that well-behaved systems of this kind can be studied by means of the theory of dominant-diagonal matrices and that the theory of dominant-diagonal matrices with finitely many elements extends in a satisfactory way to denumerable matrices. The theory of denumerable dominant diagonal matrices allows an elegant analysis of systems of intergenerational benevolence. We also revisit and extend the theory of two-sided altruism as formulated by Kimball and by Hori and Kanaya.

1. Introduction

Imagine a society in which people gain pleasure not only from their own consumption of commodities, but also from observing the happiness of others. In such a society, if Persons A, B, and C care about each other's happiness, then in order to determine how happy Person A is, one would have to know how happy B and C are. But the happiness of B and of C will depend in part on the happiness of Person A. If utility functions are to be used to represent happiness in this society, these functions must constitute a system of simultaneous equations in which the utility of each individual depends not only on his or her own private consumption, but also on the utility of others. For analysis of economic or political decision making, knowledge of individual preferences over allocations of commod-

Theodore C. Bergstrom, Department of Economics, University of California at Santa Barbara, 2127 North Hall, Santa Barbara, CA 93106-9210.
Received 23 March 1998; Accepted 27 August 1998.

Journal of Public Economic Theory, 1 (1), 1999, pp. 71–100.

ities across society is likely to be more directly useful than knowledge of individual preferences over allocations of happiness. Therefore, if we begin with knowledge about people's willingness to trade their own consumption for the happiness of others, it is useful to be able to disentangle webs of interdependent utilities to determine each individual's preferences over alternative allocations. In this paper, we show conditions under which interdependent benevolent utility functions uniquely induce well-behaved independent utility functions over consumption allocations.

2. Interdependent Utility Functions

We formalize the notion of a system of interdependent utility functions as follows. Let $\mathcal{S}$ be the set of all individuals in the society, let $\mathcal{C}_i$ be the consumption set of individual i (which is to be interpreted as the set of consumption bundles that i could conceivably consume) and let $\mathcal{C}$ be the Cartesian product $\times_{i\in\mathcal{S}}\mathcal{C}_i$. An element $c \in \mathcal{C}$ is called an *allocation*. Suppose that each individual j's *apparent happiness* is measured by a real number V_j and that preferences of any individual i depend on i's own consumption as well as on the apparent happiness of other persons in the society. For each $i \in S$, let $\mathcal{S}_{\sim i}$ be the set of all persons in S other than i and let $V_{\sim i}$ be a vector of real numbers representing the apparent happiness of each person in $S_{\sim i}$.

Let us assume that each individual i has a *private subutility function* $u_i\colon C_i \to \Re$ and an *interdependent utility function* $F_i\colon \Re^{|\mathcal{S}|} \to \Re$ such that i's preferences over i's own consumption and the apparent happiness of others are represented by

$$F_i(u_i(c_i), V_{\sim i}). \tag{1}$$

To complete the theory, we must postulate an explicit relationship between an individual's utility U_i and his or her apparent happiness V_i. In this paper we study the simplest nontrivial theory of this kind. We assume that perceived happinesses V_j accurately mirror preferences so that for all i, $V_{\sim i} = U_{\sim i}$.[1] Then for any allocation of consumption it must be that utilities satisfy the equation

$$U_i = F_i(u_i(c_i), U_{\sim i}). \tag{2}$$

We are interested in whether a system of interdependent preferences determined by private subutility functions $u_i(\cdot)$ and interdependent utility functions $F_i(\cdot,\cdot)$ determines a corresponding system of independent

[1]Several interesting alternative theories come to mind. We could introduce a dynamic theory in which at any time t, each individual i's $V_i(t)$ is equal to i's utility $U_i(t-1)$ in the previous period. We might alternatively have an expected utility theory with asymmetric information about the happiness of others, in which each individual's perception of another's "happiness" is an estimator.

utility functions $G_i: C \to \Re$ representing preferences over *allocations* for each i. Given independent utility functions $G_i(c)$ for each $i \in S$, it is convenient to define a vector-valued function $G: \mathcal{C} \to \Re^{|S|}$ such that the ith component of $G(c)$ is $G_i(c)$ and to define $G_{\sim i}(c)$ to be the vector consisting of $G_j(c)$ for all $j \in S_{\sim i}$. With this notation, we can state what it means for a system of interdependent utility functions to induce an equivalent system of independent utility functions over allocations.

Definition 1: *A system of private subutility functions $u_i: \mathcal{C}_i \to \Re$ and interdependent utility functions $F_i: \Re \times S$ will be said to induce a system of independent utility functions $G_i: \mathcal{C} \to \Re$ over allocations if for all allocations $c \in \mathcal{C}$ and for all individuals $i \in S$,*

$$G_i(c) = F_i(u_i(c_i), G_{\sim i}(c))$$

We focus our discussion by asking three questions about the relation between interdependent utility functions and induced preferences over allocations.

QUESTION 1: *When can a system of interdependent utility functions of the form described in equation (2) be disentangled to induce exactly one system of independent utility functions over allocations?*

QUESTION 2: *Given that a system of interdependent utility functions described by equation (2) induces independent utility functions $G_i: \mathcal{C} \to \Re$ for each $i \in S$, when is it true that each G_i can be written as*

$$G_i(c) = G_i(u_i(c_i), u_{\sim i}(c_{\sim i})), \tag{3}$$

where $u_{\sim i}(c_{\sim i})$ is the vector whose components are the private subutility functions $u_j(c_j)$ for $j \neq i$ and where G_i is a nondecreasing function of u_j for all $j \in S$?

It is useful to define *normal benevolence* in such a way that Questions 1 and 2 can be expressed simply by asking "When is a system of interdependent utilities *normally benevolent*?"

Definition 2: *A system of interdependent utilities with private subutility functions $u_i(c_i)$ and interdependent utility functions $U_i = F_i(u_i(c_i), U_{\sim i})$* is *normally benevolent if it induces a unique system of independent utility functions $G_i: \mathcal{C}_i \to \Re$ for $i \in \mathcal{S}$ such that $G_i(c) = G_i(u_i(c_i), u_{\sim i}(c_{\sim i}))$, where each G_i is a monotone increasing function of each of the u_j*'s.

Although in a theory of interdependent utility functions individuals must necessarily make interpersonal comparisons of utility, the theory proposed here remains ordinalist in the sense that the same system of ordinal preferences over allocations can be represented by more than one system of interdependent utility functions—so long as these functions are related to each other by an appropriate class of transformations. Of course when utility functions are interdependent, we cannot expect preferences over allocations to remain unaltered if we simply rescale the utility function of one individual without making corresponding adjustments in the way that the rescaled utility function enters the interdependent utility functions of those who care about the happiness of the person with rescaled utility. The following definition will be useful:

DEFINITION 3: *A system of interdependent utility functions with private subutility functions u_i for each i and interdependent utility functions $F_i(u_i(c_i), V_{\sim i})$ is said to be equivalent to another system of interdependent utility functions defined by private subutility functions $\tilde{u}_i(c_i)$ for each i and by interdependent utility functions $\widetilde{F}_i(\tilde{u}_i(c_i), V_{\sim i})$ if both systems induce the same preferences over allocations for all i.*

QUESTION 3: *When are two different systems of interdependent utility functions equivalent?*

2.1 A Two-Person Example

We can take an instructive first pass at Questions 1–3 by looking at the case of two consumers, Romeo and Juliet.[2] Romeo and Juliet each consume a single commodity, spaghetti. By staring deeply into one anothers' eyes, each can determine the other's happiness. Each finds happiness from consuming spaghetti and from observing the other's happiness. Their interdependent utility functions take the additively separable form:

$$U_R = F_R(u_R(c_R), U_J) = u_R(c_R) + aU_J \tag{4}$$

$$U_J = F_J(u_J(c_J), U_R) = u_J(c_J) + bU_R, \tag{5}$$

where c_R and c_J are, respectively, Romeo's and Juliet's consumption of spaghetti and where a and b are positive constants.

If $ab \neq 1$, one can eliminate U_R and U_J from the right-hand sides of equations (4) and (5) and solve uniquely for utility functions $G_R(u_R(c_R), u_J(c_J))$ and $G_J(u_R(c_R), u_J(c_J))$ that are defined on allocations. In particular,

$$U_R = G_R(u_R(c_R), u_J(c_J)) = \left(\frac{1}{1-ab}\right) u_R(c_R) + \left(\frac{a}{1-ab}\right) u_J(c_J) \tag{6}$$

$$U_J = G_J(u_R(c_R), u_J(c_J)) = \left(\frac{1}{1-ab}\right) u_J(c_J) + \left(\frac{b}{1-ab}\right) u_R(c_R). \tag{7}$$

For this example, the answer to our earlier Question 1 is immediate from equations (6) and (7). The system of interdependent utility functions (4) and (5) induces a unique system of independent utility functions given by equations (6) and (7) over allocations if and only if $ab \neq 1$.

[2]This case is also treated in Bergstrom (1989).

The answer to Question 2 for this example is also immediate from equations(6) and (7), which show that G_R and G_J are both nondecreasing functions of u_R and u_J if and only if $a > 0$, $b > 0$, and $ab < 1$.

If $a > 0$, $b > 0$, and $ab > 1$, the functions G_R and G_J are seen to be monotone *decreasing* in their arguments. At first glance, this result seems puzzling because if $u_R(\cdot)$ and $u_J(\cdot)$ are increasing functions of c_R and c_J, this would imply that in the independent utility functions over allocations expressed by equations (6) and (7), both Romeo and Juliet prefer less spaghetti to more. The apparent paradox is resolved by the observation that if the interdependence of their utilities is so strong that $ab > 1$, then whenever Romeo and Juliet disagree about allocations of spaghetti it is because each wants the other to take the larger portion. Where $ab < 1$, both persons can have independent preferences in which they prefer more spaghetti to less if and only if $u_R(\cdot)$ and $u_J(\cdot)$ are *decreasing* functions of c_R and c_J. In this case, although he would prefer more spaghetti to less when holding Juliet's *spaghetti consumption* constant, Romeo would prefer less spaghetti to more for himself if Juliet's *utility* must be held constant during the experiment, because the only way to hold Juliet's utility constant while increasing Romeo's spaghetti consumption is to reduce Juliet's spaghetti consumption by so much as to make Romeo less happy.

Suppose that the private subutility functions for Romeo and Juliet are recalibrated to take on new values,

$$\tilde{u}_R(c_R) = d_R u_R(c_R) \quad \text{and} \quad \tilde{u}_J(c_J) = d_J\, u_J(c_J), \tag{8}$$

respectively, where $d_R > 0$ and $d_J > 0$. With the recalibrated private subutility functions we can maintain the original preferences over allocations only if we adjust the responses of each person's utility to the other's recalibrated utility so as to take into account the change in units of measurement. Specifically, let

$$\tilde{a} = \frac{d_R}{d_J}\, a \quad \text{and} \quad \tilde{b} = \frac{d_J}{d_R}\, b \tag{9}$$

and define adjusted interdependent utility functions $\tilde{F}_R(\cdot,\cdot)$ and $\tilde{F}_J(\cdot,\cdot)$ so that:

$$\tilde{F}_R(\tilde{u}_R(c_R), U_j) = \tilde{u}_R(c_R) + aU_J \tag{10}$$

$$\tilde{F}_J(\tilde{u}_J(c_J), U_R) = \tilde{u}_J(c_J) + \tilde{b}U_R. \tag{11}$$

The interdependent utility functions (10) and (11) induce independent utility functions over allocations $\tilde{G}_R(\cdot,\cdot)$ and $\tilde{G}_J(\cdot,\cdot)$, where

$$\tilde{G}_R(\tilde{u}_R(c_R), \tilde{u}_J(c_J)) = \left(\frac{1}{1-\tilde{a}\tilde{b}}\right)\tilde{u}_R(c_R) + \left(\frac{\tilde{a}}{1-\tilde{a}\tilde{b}}\right)\tilde{u}_J(c_J) \tag{12}$$

$$\tilde{G}_J(\tilde{u}_R(c_R), \tilde{u}_J(c_J)) = \left(\frac{1}{1-\tilde{a}\tilde{b}}\right)\tilde{u}_J(c_J) + \left(\frac{\tilde{b}}{1-\tilde{a}\tilde{b}}\right)\tilde{u}_R(c_R). \tag{13}$$

Substituting from equations (8) and (9) into equations (12) and (13) and noticing that $ab = \tilde{a}\tilde{b}$, we find that the latter two equations are equivalent to

$$\tilde{G}_R(\tilde{u}_R(c_R), \tilde{u}_J(c_J)) = d_R G_R(u_R(c_R), u_J(c_J)) \tag{14}$$

$$\tilde{G}_J(\tilde{u}_R(c_R), \tilde{u}_J(c_J)) = d_J\, G_J(u_R(c_R), u_J(c_J).). \tag{15}$$

Since the induced independent utility functions $\tilde{G}_R(\cdot,\cdot)$ and $\tilde{G}_J(\cdot,\cdot)$ are monotone transformations of the utility functions $G_R(\cdot,\cdot)$ and $G_J(\cdot,\cdot)$, it follows that the two interdependent systems of utility functions are equivalent.

3. Benevolent Interaction with Dominant Diagonals

In the example of Romeo and Juliet, we solve a pair of simultaneous equations to find the unique independent system of utility functions induced by the system of interdependent utility functions (4) and (5). More generally, we will show that if preferences are benevolent, but not too intensely benevolent, induced independent utility functions can be found by inverting a dominant diagonal matrix.

Interdependent preferences are likely to be particularly strong between family members. Moreover, intrafamilial utility interdependence often has an interesting special structure. For example, suppose that parents care about the happiness of their children and children care about the happiness of their parents. Since the happiness of each generation depends in part on that of its predecessor and in part on that of its successor, the generations are indirectly linked far into the distant future and the distant past. In order to treat intergenerational preferences in an elegant and convenient way, it is useful to deal with a doubly infinite sequence of generations—where we consider ancestors running back to an infinite past and descendants running forward to an infinite future. Even though we may believe that the true model is one with a long but finite past and a long but finite future, approximation by an infinite horizon model is attractive. This procedure allows us to state crisp results that exclude the complicated but inessential details that arise if each succeeding generation is a little closer to the end of the world and a little further from the beginning of the world.

Approximation of models with distant finite horizons by infinite-horizon models is only appropriate if consumption differences in the sufficiently distant future and past have arbitrarily small influence on preferences between two time paths of consumption. Dominant diagonal matrices with a denumerable infinity of rows and columns are shown to be a useful mathematical tool for analysis of systems of benevolent utility functions extending to the distant past and distant future. Fortunately it turns out that the powerful and well-known theory of dominant-diagonal matrices with finitely many entries extends cleanly to denumerable dominant diagonal matrices.

3.1 Additively Separable Interdependent Utility Functions

We begin by considering additively separable systems of interdependent utility functions. These utility functions can be written as follows:

$$U_i = F_i(u_i(c_i), V_{\sim i}) = u_i(c_i) + \sum_{j \neq i} a_{ij} V_j, \tag{16}$$

where $a_{ij} \geq 0$ for all i and j. For our subsequent discussion it is useful to define the following vector-valued functions.

Definition 4: *Define the function $u: \mathcal{C} \to \Re^{|S|}$ so that $u(c)$ is the vector whose ith element is $u_i(c_i)$ for each $i \in \mathcal{S}$. Define the function $F: \Re^{|S|} \times \Re^{|S|}$ so that $F(u(c), V_{\sim i})$ is the vector whose ith element is $F_i(u_i(c_i), V_{\sim i})$ for each $i \in \mathcal{S}$.*

Where we identify the utility vectors U with the vectors of apparent happiness V, this system of interdependent utility functions is described by the matrix equation

$$U = F(u(c), U) = u(c) + AU. \tag{17}$$

When $(I - A)^{-1}$ exists, the system (17) induces an independent system of utility functions represented by the matrix equation

$$U = (I - A)^{-1} u(c). \tag{18}$$

Therefore the system (16) is normally benevolent if and only if the matrix $(I - A)^{-1}$ exists and is nonnegative in every element.

3.2 Dominant Diagonal Matrices, Finite and Infinite

The theory of finite-dimensional dominant diagonal matrices is a familiar tool for economists.[3] A matrix that is of the form $I - A$ where $A > 0$ is dominant diagonal if its row sums are all positive or if it can be postmultiplied by a diagonal matrix to produce a matrix with positive row sums. If $I - A$ is a finite-dimensional dominant diagonal matrix, then it turns out that $(I - A)^{-1}$ exists and is nonnegative in every entry.

To deal with intergenerational models, we need to extend this result, insofar as possible, to denumerably infinite dominant diagonal matrices. The notation of linear algebra and its operations extends in the obvious way to denumerably infinite matrices and vectors. Though many of the

[3] See McKenzie (1959) or Gale (1960), pp. 294–301).

fundamental results of finite dimensional linear algebra carry over to this environment, there are some nasty surprises. Among these surprises are the fact that matrix multiplication is not generally associative and the fact that a matrix may have more than one inverse.[4] Fortunately, denumerable dominant diagonal matrices are much better behaved than denumerable matrices in general.

For our analysis, a vector x has a denumerable infinity of components, x_i, where the index i runs from $-\infty$ to ∞. A matrix M has components M_{ij} where i and j run from $-\infty$ to ∞. Matrix multiplication is defined in exact analogy to the case of finite matrices, with the elements of the product being the appropriate sums of an infinite series. A matrix is D is said to be a *diagonal* matrix if the only nonzero elements of D are the diagonal elements D_{ii}. A matrix M is said to be *bounded* if there exists some real number b such that $|M_{ij}| \leq b$ for all i and j. Obviously, every finite matrix is bounded, but a denumerable matrix might not be. (For example, the denumerable matrix in which $M_{ij} = i$ for all positive integers i and j is not bounded.)

DEFINITION 5: *A denumerable matrix $I - A$ such that $A \geq 0$ is said to be dominant diagonal if there exists a bounded diagonal matrix $D \geq 0$ such that the infimum of the row sums of the matrix $(I - A)D$ is positive.*

From elementary matrix calculations it can be seen that an equivalent definition of a denumerable dominant diagonal matrix is as follows:

Remark 1: A denumerable matrix $I - A$ such that $A \geq 0$ is dominant diagonal if and only if there exists a bounded vector $y > 0$ such that $(I - A)y >> 0$.

Evidently a matrix M is dominant diagonal if each row sum of M is positive. Although this condition is sufficient, it is not necessary. For example, consider the two-by-two matrix

$$M = \begin{pmatrix} 1 & -\frac{3}{2} \\ -\frac{1}{3} & 1 \end{pmatrix}.$$

Although the row sums of M are not all positive, there exists a diagonal matrix D such that the row sums of MD are positive. For example, let $D_{11} = 1$ and $D_{22} = 1/2$. Then

$$MD = \begin{pmatrix} 1 & -\frac{3}{2} \\ -\frac{1}{3} & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & -\frac{3}{4} \\ -\frac{1}{3} & \frac{1}{2} \end{pmatrix}$$

has positive row sums and therefore M must be a dominant diagonal matrix.

[4]For a good exposition of this theory, see Kemeny, Snell, and Knapp (1966).

Although a denumerable dominant diagonal matrix in general has more than one inverse, it has only one *bounded* inverse matrix. Though it is possible to define utility functions of the form $U = Bu$ where B is unbounded, the preference relations they induce are peculiar. The issues surrounding unbounded utility representations are treated by Koopmans (1972). Indeed, if B is unbounded, the use of a model with infinite horizons as as an approximation to a finite horizon model becomes untenable.

Our principal mathematical result concerning dominant diagonal matrices is the following lemma, which is proved in the Appendix.[5]

LEMMA 1: *Let $I - A$ be a finite or denumerable matrix such that $A \geq 0$. Then $I - A$ has a unique bounded inverse matrix $B \geq 0$ if and only if $I - A$ is dominant diagonal. This inverse matrix is $B = \sum_{t=0}^{\infty} A^t$.*

3.3 Regular, Normally Benevolent Preferences

Let us define preferences over allocations to be *regular* if they can be represented by a utility function that takes a finite value for every constant allocation (…, c, c, …). Clearly, preferences that can be represented by a utility function of the form $U = Bu$ are *regular* if and only if B is bounded. As a consequence of Lemma 1 we have the following result.

PROPOSITION 1: *If $A \geq 0$ and if $I - A$* is *dominant diagonal, then exactly one set of regular preferences over allocations is induced by the interdependent utility system $U = F(u(c), U) = u(c) + AU$.*

The following proposition, which follows directly from Lemma 1, is an answer to our earlier Questions 1 and 2 when utility functions are additively separable.

PROPOSITION 2: *A finite or denumerably infinite system of interrelated utility functions of the form $U_i = u_i(c_i) + \sum_{j \neq i} a_{ij} U_j$ with $a_{ij} \geq 0$ for all i and j is normally benevolent if $I - A$ is a dominant diagonal matrix where A is the matrix whose ijth entry is a_{ij}.*

3.4 Nonadditive Interdependent Utilities

The results of Proposition 2 extend to weakly separable utility functions that are continuously differentiable but not necessarily additively separable.

PROPOSITION 3: *Consider a system of interdependent utility functions described by the equations $U_i = F_i(u_i, U_{\sim i})$ for all individuals i. Let $A(u_1(c_1), \ldots, u_n(c_n))$ be the matrix of partial derivatives whose ijth element is the derivative of F_i with respect to U_j. If $I - A(u_1(c_1), \ldots, u_i(c_i), \ldots)$ is dominant diagonal, then the system*

[5] The proof is a straightforward extension of a standard proof of this proposition for finite matrices.

of utility functions is normally benevolent in an open neighborhood of $(c_1, \ldots, c_i, \ldots)$. *Moreover, if* $I - A(u_1(c_1), \ldots, u_i(c_i), \ldots)$ *is dominant diagonal at all allocations, then preferences are normally benevolent everywhere.*

The first assertion of Proposition 3 is immediate from the implicit function theorem. The second assertion follows from the Gale–Nikaido theorem on the existence of global inverses for nonlinear functions on convex sets (Nikaido 1968). The conditions we impose on the linear function imply that the Jacobian of the vector-valued function $U - F(u, U)$, treated as a function of the vector of utilities U, will be a P-matrix and hence the Gale–Nikaido theorem can be applied.

4. Equivalent Systems of Utility Functions

Consider two regular, normally benevolent systems of interdependent utility functions where the first system is described by private subutility functions $u_i(c_i)$ for each i and by interdependent utility functions for each i of the form

$$F_i(u_i, U_{\sim i}) = u_i(c_i) + \sum_{j \neq i} a_{ij} U_j, \tag{19}$$

and where the second system is described by private subutility functions $\tilde{u}_i(c_i)$ for each i and by interdependent utility functions of the form

$$\tilde{F}_i(\tilde{u}_i, U_{\sim i}) = \tilde{u}_i(c_i) + \sum_{j \neq i} \tilde{a}_{ij} U_j. \tag{20}$$

Let us define functions $u(\cdot)$ and $\tilde{u}(\cdot)$ so that $u(c)$ and $\tilde{u}(c)$ are, respectively, the vectors whose ith components are $u_i(c_i)$ and $\tilde{u}_i(c_i)$, and let us also define the vector-valued functions $F(\cdot,\cdot)$ and $\tilde{F}(\cdot,\cdot)$ so that the ith component of $F(u, U)$ is $F_i(u_i, U)$ and the ith component of $\tilde{F}(\tilde{u}, U)$ is $\tilde{F}_i(\tilde{u}_i, U)$. Then we can express the systems of equations (19) and (20), respectively, by the matrix expressions

$$F(u(c), U) = u(c) + AU \tag{21}$$

and

$$\tilde{F}(\tilde{u}(c), U) = \tilde{u}(c) + \tilde{A}U, \tag{22}$$

where the ijth entries of the matrices A and $\tilde{A}$ are, respectively, a_{ij} and $\tilde{a}_{ij}$.

The following proposition establishes conditions under which two systems of interdependent utilities represent the same preferences over allocations.

PROPOSITION 4: *Consider two systems of regular, normally benevolent interdependent utilities which are described by private subutility functions* $u_i(c_i)$ *and* $\tilde{u}_i(c_i)$, *respectively, for each* i *and by interdependent utility functions* $F(u(c), U) =$

$u(c) + AU$ and $\tilde{F}(\tilde{u}(c), U) = \tilde{u}(c) + \tilde{A}U$. These systems are equivalent if and only if there is a diagonal matrix D with strictly positive diagonals such that $\tilde{A} = DAD^{-1}$ and $\tilde{u} = Du + \tilde{b}$ for some vector of constants $\tilde{b}$.

Proof: The systems of interdependent utility functions given by (21) and (22) are equivalent if and only if they induce the same system of independent preferences over allocations. The independent utility functions over allocations induced by these two systems are given, respectively, by the matrix equations

$$G(c) = (I - A)^{-1} u(c) \tag{23}$$

and

$$\tilde{G}(c) = (I - \tilde{A})^{-1} u(c) \tag{24}$$

A standard result of consumer theory (see, e.g., Debreu 1960) is that if two different *additively separable* functions represent the same preferences, they must be affine transformations of each other. Therefore the two systems are equivalent if and only if for each i there are constants $d_i > 0$ and b_i such that $\tilde{G}_i(c_i) = d_i G_i(c_i) + b_i$. These conditions can be expressed simultaneously for all i in terms of matrices as $\tilde{G}(c) = DG(c) + b$, or equivalently as

$$G(c) = D^{-1}\tilde{G}(c) - D^{-1}b, \tag{25}$$

where D is the diagonal matrix for which $D_{ii} = d_i$ and where b is the vector whose ith entry is b_i. From equation (23) it follows that

$$G(c) = u(c) + AG(c). \tag{26}$$

If we substitute the expression in equation (25) for $G(c)$ on both sides of equation (26) and then premultiply both sides of the resulting equation by D and rearrange terms, we find that

$$\tilde{G}(c) = Du(c) + (I - DAD^{-1})b + DAD^{-1}\tilde{G}(c). \tag{27}$$

But we also know, as a consequence of equation (24), that

$$\tilde{G}(c) = \tilde{u}(c) + \tilde{A}\tilde{G}(c). \tag{28}$$

From equations (27) and (28) we must conclude that $\tilde{u}(c) = Du(c) + (I - DAD^{-1})b$ and that $\tilde{A} = DAD^{-1}$. ■

As a consequence of Proposition 4, standard results from linear algebra can establish several properties that must be shared by the matrices A and $\tilde{A} = DAD^{-1}$ if the two systems of interdependent utility functions described by equations (21) and (22) are equivalent.

COROLLARY 1: *Consider two systems of regular, normally benevolent interdependent utilities which are described by private subutility functions $u_i(c_i)$ and $\tilde{u}_i(c_i)$,* *respectively, for each i and by interdependent utility functions* $F(u(c), U) = u(c) + AU$ *and* $\tilde{F}(\tilde{u}(c), U) = \tilde{u}(c) + \tilde{A}U$. *If these systems are equivalent, then*

- *For all i and j, the sign of the ijth entry of* $\tilde{A}$ *is the same as the sign of the ijth entry of A.*
- *The determinant of* $\tilde{A}$ *is the same as the determinant of A.*
- *The eigenvalues of* $\tilde{A}$ *are the same as the eigenvalues of A.*
- *The matrix* $I - \tilde{A}$ *is dominant diagonal if and only if the matrix* $I - A$ *is dominant diagonal.*

5. Two-Sided Intergenerational Altruism

An overlapping generations model in which each generation cares about its own consumption and about the utility of the next generation is a simple example of a system of interrelated utility functions. In this model, preferences are assumed to be stationary across generations and the utility function of the tth generation is assumed to be

$$U_t = u(c_t) + bU_{t+1}, \tag{29}$$

where $0 < b < 1$ (see, for example, Barro 1974). It is well known that this system of interdependent utility functions induces independent utility functions for generation t which take the form

$$G_t(c) = u(c_t) + \sum_{s=1}^{\infty} b^s u(c_{t+s}). \tag{30}$$

A more complicated and interesting structure arises if each generation cares not only about its own consumption and the utility of its successor, but each generation also cares about the utility of its parent generation. Kimball (1987) gave this preference structure the name "two-sided altruism." This problem, which perplexed early researchers, has a beautiful solution, which was discovered independently by Kimball and by Hori and Kanaya (1989). Kimball and Hori and Kanaya exploited the special structure of the two-sided altruism problem by posing the mapping from interdependent utility functions to independent utilities over allocations as a solution to a system of stationary difference equations.

This paper offers an alternative method of solving the problem of two-sided altruism, using the theory of dominant diagonal matrices. An advantage of the matrix approach is that the qualitative results extend directly to nonstationary cases and to utility interdependence among arbitrary sets of consumers.[6] Additionally, this presentation attempts to clarify

[6]Hori (1992) studied the case of stationary intergenerational utility interdependence that extends for an arbitrary number of generations in either direction. For this general case, he was not able to find the utility functions over allocations in explicit form, but he displays a system of simultaneous equations that must be satisfied by the parameters of these utilities.

the exposition and comment on some puzzling issues raised by the earlier authors.

Suppose that preferences are additively separable and stationary over time. The utility function of generation t is assumed to take the form

$$U_t = u(c_t) + aU_{t-1} + bU_{t+1}. \tag{31}$$

The system of interrelated utilities represented in equation (31) invites alternative interpretations that may be at least as interesting as the intergenerational model. For example, this model would apply directly to a population of individuals living along a road, each of whom is concerned about his own consumption and that of his immediate neighbors on either side. Perhaps even more intriguing is to think of the two-sided altruism as a simple model of memory and anticipation for a single person. Thus equation (31) could determine the time path of happiness for an individual whose happiness at time t depends on what she consumes at time t, on her remembered happiness from the previous period, and on her anticipated happiness in the next period.[7]

Kimball (1987) suggests a nice metaphor for the backward and forward reflections of utility in the case of two-sided altruism, which he calls a "Hall of Mirrors effect." In the intergenerational interpretation, part of a child's happiness comes from observation of her parent's happiness. The parent's happiness is in turn influenced both by the happiness of the child and by the happiness of the grandparent, and so on, reflecting across the generations. This metaphor is even more striking if we think about the interpretation of two-sided altruism as a dynamic model of individual happiness, incorporating memory and anticipation. A person's current sense of well-being is formed by reflections through corridors of remembered pleasures and pains and of contemplated future delights and torments. For the two-sided altruism model, where $a + b < 1$, although current events bear intimations of the remote past and distant future, these influences turn out to be damped and the dynamic path of utility is well defined and stable.

Equation (31) can be expressed as a matrix equation $U = u + AU$, or equivalently by $U = (I - A)u$, where U and u are vectors whose tth components are U_t and $u_t(c_t)$, respectively, and where A is a matrix with values a everywhere on the first subdiagonal, b everywhere on the first superdiagonal, and zeros everywhere else. The matrix $I - A$ is dominant diagonal if and only if $a + b < 1$ and, hence, according to Proposition 2 the system of interdependent utilities is normally benevolent if $a + b < 1$. Since $I - A$ is dominant diagonal, it has a unique bounded inverse and we can therefore find the induced system of independent utility func-

[7]For a more satisfactory two-sided altruism model of memory and anticipation, it would be useful to explicitly incorporate nonstationarity over time and to recognize the finiteness of individual life. In addition to other differences, young people are different from old people because they have a shorter past and a longer future.

tions over allocations as $U = (I - A)^{-1}u$, where $(I - A)^{-1}$ is the unique bounded inverse of the matrix $I - A$. Where B_{tj} is the entry in the tth row and jth column of $I - A$, the preferences of individual t over allocations can then be represented by the utility function $G_t(c) = \sum_{j=-\infty}^{\infty} B_{tj}\, u(c_j)$. In the Appendix, we solve for the matrix $B = (I - A)^{-1}$ and thus find an explicit expression for the independent utility function $G_t(c)$.

PROPOSITION 5: *Let there be a denumerable infinity of consumers with interdependent utility functions of the form*

$$U_t = u(c_t) + aU_{t-1} + bU_{t+1} \tag{32}$$

for every generation t, with $a \geq 0$ *and* $b \geq 0$. *This system is normally benevolent if and only if* $a + b < 1$. *If* $a + b < 1$, *then the preferences of the tth generation over allocations can be represented by a utility function of the form*

$$G_t(c) = \sum_{j=1}^{\infty} \alpha^j u(c_{t-j}) + u(c_t) + \sum_{j=1}^{\infty} \beta^j u(c_{t+j}), \tag{33}$$

where $0 < \alpha < 1,\ 0 < \beta < 1,$ *and where*

$$\alpha = \frac{1 - \sqrt{1 - 4ab}}{2b} \quad \text{and} \quad \beta = \frac{1 - \sqrt{1 - 4ab}}{2a}. \tag{34}$$

The preferences represented by these utility functions are the only regular *preferences over allocations that are consistent with the original system of interrelated utility functions.*

According to Proposition 5, in the case of two-sided altruism with $a + b < 1$, each individual's utility function over allocations is a weighted sum of the private subutilities $u_i(c_i)$, where an individual gives descendants and ancestors weights that decrease exponentially as the relation becomes more distant. These discount rates α and β are jointly determined by the parameters a and b of the interdependent utilities in expression (32). Lemma 2 shows that the mapping from the parameters a and b to the discount rates α and β is one-to-one. Lemma 3 gives us useful information about the way that the weights a and b in the interdependent utilities determine the rates α and β at which the past and future are discounted in the independent utility functions.

LEMMA 2: *Consider the mapping F with domain* $\mathcal{X} = \{(a, b) | a \geq 0,\ b \geq 0,\ a + b < 1\}$, *and range* $\mathcal{Y} = \{(\alpha, \beta) | 0 \leq \alpha < 1,\ 0 \leq \beta < 1\}$ *such that*

$$F(a, b) = (\alpha(a, b), \beta(a, b)),$$

and where if $a > 0$ *and* $b > 0$,

$$\alpha(a, b) = \frac{1 - \sqrt{1 - 4ab}}{2b} \quad \text{and} \quad \beta(a, b) = \frac{1 - \sqrt{1 - 4ab}}{2a},$$

and where $\alpha(a, 0) = a$ *and* $\beta(0, b) = b$.

- *The mapping F is a continuous bijection (one-to-one and onto) from the set $\mathcal{X}$ onto the set $\mathcal{Y}$.*
- *The inverse mapping is defined by* $F^{-1}(\alpha,\beta) = (\alpha/(1+\alpha\beta),(\beta/(1+\alpha\beta))$.

LEMMA 3: *The mapping* $F(a,b) = (\alpha(a,b),\beta(a,b))$ *defined in Lemma 2 has the following properties for all* $(a,b) \in \mathcal{X}$.

- $\dfrac{\alpha(a,b)}{\beta(a,b)} = \dfrac{a}{b}$
- *The partial derivatives of* $\alpha(a,b)$ *and* $\beta(a,b)$ *with respect to a and b are positive for all positive a and b such that* $a + b < 1$.

From Lemma 3, we see that the ratio of discount rates assigned to consumption of past generations and future generations in the independent utility functions is the same as the ratio of the weights a and b. Perhaps more surprisingly, we also see that the rate at which the current generation values the consumption of future generations is an increasing function not only of the weight b that individuals attach to the utility of their children, but also of the weight a that individuals attach to the utility of their parents.

The next lemma shows that the matrix $I - A$ in the two-sided altruism model can be decomposed into the product of a matrix with zeroes everywhere above the diagonal times another matrix with zeroes everywhere below the diagonal. This fact makes it easy to determine the inverse $(I - A)^{-1}$ and also enables us to decompose the forward-looking and backward-looking sides of two-sided altruism in an informative way.

LEMMA 4: *Let* $I - A = (I - aJ^{-1} - bJ)$, *where* J^{-1} *is the matrix with* 1*'s on the first subdiagonal and* 0*'s everywhere else, and where J is the matrix with* 1*'s on its first superdiagonal and* 0*'s everywhere else, and where* $a > 0$, $b > 0$, *and* $a + b < 1$. *Let* $\alpha = \alpha(a,b)$ *and* $\beta = \beta(a,b)$ *as defined by equations (34). Then*

$$I - A = \left(\frac{1}{1+\alpha\beta}\right)(I - \alpha J^{-1})(I - \beta J). \tag{35}$$

6. Agreement and Conflict between Generations

In this section, we address two questions that arise in interpreting the two-sided altruism model and related models of altruism across generations.

QUESTION 4: *In models of benevolent interaction between generations (or between present and future selves) what will different generations (or different selves) agree about and what will they disagree about?*

If we interpret the two-sided altruism model as a spatial model of interaction between neighbors located on a line, there is no special reason to distinguish between persons located to one's left and persons located to one's right. But if we interpret two-sided altruism as a model of interaction between neighbors in time rather than in space, there may be important reasons to treat the past differently from the future. This leads to our next question.

QUESTION 5: *If we believe that time is irreversible, that consumption in the past cannot be altered by current decisions, and that one's dead ancestors are not conscious of our current actions, how should we build these views into our modeling of interdependent utilities across generations?*

6.1 Forward and Backward Consistency

Strotz (1955) posed the question of intertemporal consistency of choice as follows.

> An individual is imagined to choose a plan of consumption for a future period of time so as to maximize the utility of the plan as evaluated at the present moment. . . . If he is free to reconsider his plan at later dates, will he abide by it or disobey it—*even though his original expectations of future desires and means of consumption are verified*?

This question applies with equal force to intertemporal planning for an individual and to intergenerational allocation for a benevolently linked family. This problem is well understood for the case of one-sided altruism, where individuals care about the utility of their children, but not about their parents. The one-sided altruism model also applies to individuals whose current happiness depends on their current consumption and anticipated future happiness, but not on their recollected happiness.[8] Koopmans (1960) demonstrates that if preferences over allocations of consumption across time (or across generations) are continuous, additively separable across time periods, and stationary over time, then they will induce time-consistent choices in the sense proposed by Strotz if and only if they can be represented by utility functions that take the form found in equation (30) where consumption in future periods is discounted at a constant per-period rate.

The older literature on time-consistency dealt only with "forward-looking" consistency in models of one-sided altruism. When there is two-sided altruism between generations, it becomes important to consider questions related to "backward-looking" consistency between generations as well. These issues were raised in an interesting debate between Buiter and Carmichael (1984) and Burbidge (1983, 1984), which was motivated by the study of social security systems, the support for the elderly by their

[8]Several authors have followed Strotz in proposing and studying interesting models of intertemporal decision making in the absence of consistent preferences. A good discussion of this issue can be found in Blackorby et al. (1973).

children, and the impact of national debt. Buiter and Carmichael proposed a utility function for consumption across generations of the form

$$U_t = \alpha u(c_{t-1}) + u(c_t) + \sum_{s=1}^{\infty} \beta^s u(c_s). \tag{36}$$

Burbidge (1983) pointed out that with the Buiter-Carmichael utility function, in order for each generation to agree with its parent generation about allocations of resources that differ only with respect to consumption of the parent's generation and later generations, then it must be that $\alpha = 1/\beta$. As Burbidge points out, this utility function has the odd characteristic that individuals "reverse-discount" their parent's consumption, giving it more weight than they give to their own.

In contrast to Burbidge's result, the two-sided altruism model induces a utility function in which each generation's utility function gives a lower weight to its parent's consumption than it does to its own, and also discounts consumption of all preceding generations, using weights that decrease exponentially with distance in time. The reason for this difference is that the two-sided altruism model demands less agreement between parent and child than Burbidge proposes. In both models, a parent and child must agree about allocations that differ only in the consumption of the child and her descendants. However, the two-sided altruism model, unlike the Burbidge-Buiter-Carmichael model, allows the possibility that parent and child may have conflicting views about how to divide consumption between themselves.[9]

Let us define there to be *forward-looking consistency of preferences between adjacent generations* if every generation t agrees with its successor in its preference ranking of alternative consumption bundles that differ only in the allocation of consumption among individuals belonging to generation $t + 1$ or later and *backward-looking consistency of preferences between adjacent generations* if every generation t agrees with its predecessor in its preference ranking of alternative consumption bundles that differ only in the allocation of consumption among individuals belonging to generation $t - 1$ or earlier. The next result, which is proved in the Appendix, is a simple consequence of the form of the utility functions over allocations reported in Proposition 5. This provides us with a partial answer to Question 4.

PROPOSITION 6: *In the model of two-sided altruism where $a + b < 1$, there is forward-looking and also backward-looking consistency of preferences between adjacent generations.*[10]

[9] The question of which of the two models is closer to reality is an empirical one, though I suspect that the evidence favors the view that children weight their parents' consumption less rather than more than their own.

[10] Although it is not possible, given current technology, to go back and rearrange the past, it might interest historians to notice that with two-sided altruism of this type, all generations agree with their predecessors about what they *would have liked* their history to have been. If this were the case, there would be little work for historical revisionists.

6.2 The Will of Dead Ancestors

Within the framework of the two-sided altruism model, Hori and Kanaya (1989) propose two alternative ways of treating the interests of individuals who are no longer alive.

> One natural way would be to regard past generations' utility levels as givens. If the framework is that of non-overlapping generations where each generation lives only one period, then U_{t-1} is given when the tth generation makes the choice. (p. 245)

Hori and Kanaya propose that if this is the viewpoint taken, then for the two-sided altruism model specified by equation (31), the induced utility function of interest takes the following form:[11]

$$U_t = H(U_{t-1}, {}_t u(c)), \tag{37}$$

where ${}_t u(c)$ denotes the vector of private subutilities $(u_t(c_t), u_{t+1}(c_{t+1}), \ldots)$ for consumers of generation t or later. Hori and Kanaya find the unique function of this form that is consistent with equation (31) and they call this function a *mortality solution.*

Hori and Kanaya (1989) also consider an alternative view, which they justify in the following way.

> But this is not the only way to treat past generations. Each individual might possibly be so obedient that he may care how his parents would respond to his and his descendants' future consumption plans if they were alive. In this case, parents bodies die but their souls live eternally and it is no longer possible to regard past generations' utilities as givens. (p. 245)

Hori and Kanaya propose that the independent utility function $G(c)$ specified in Proposition 5 is the appropriate utility function if this is the way people regard the will of their dead ancestors. Hori and Kanaya call this function an *eternity solution.*

We show in the Appendix that Hori and Kanaya's mortality solution can be found in a straightforward way using matrix methods and the product decomposition found in Lemma 4. There are at least two notable features of the mortality solution. First, we see that the domain of values of parameters a and b for which there is a bounded mortality solution is larger than that for which there are bounded utility functions defined over the infinite past as well as the infinite future. This happens because in order for the mortality solution to be bounded, we do not need $\alpha(a, b) < 1$, but only $0 \leq \beta < 1$. Second, we see that when $a + b < 1$, for all $s > 0$, the weight β^s that generation t puts on the subutility function for generation $t + s$ is the same as that which appears in the independent

[11] I have slightly changed the notation from that used by Hori and Kanaya to be consistent with that used in the rest of this paper.

utility function G_t in equation (33) (which Hori and Kanaya call the eternity solution). It may seem surprising that even holding the utility of the preceding generation constant, the discount rate $\beta(a, b)$ that generation t uses to weigh the subutilities of one successor relative to another incorporates the full effect of the weight a that each generation attaches to the utility of its predecessor. Stated in our notation, Hori and Kanaya's mortality solution is as follows.

PROPOSITION 7: *In the two-sided altruism model, if* $a + b < 1$ *or if* $a + b > 1$ *and* $b > 1/2$ *then equation (31) describes the preferences of each generation t if and only if preferences of each individual t can also be represented by the bounded utility function*

$$U_t = H(U_{t-1}, {}_t c) = \alpha U_{t-1} + u(c_t) + \sum_{s=1}^{\infty} \beta^s u(c_{t+s}), \tag{38}$$

where

$$\alpha = \frac{1 - \sqrt{1 - 4ab}}{2b} \quad \text{and} \quad \beta = \frac{1 - \sqrt{1 - 4ab}}{2a} \tag{39}$$

if $a + b < 1$ *and where*

$$\alpha = \frac{1 + \sqrt{1 - 4ab}}{2b} \quad \text{and} \quad \beta = \frac{1 + \sqrt{1 - 4ab}}{2a} \tag{40}$$

if $a + b > 1$ *and* $b < 1/2$.

Although Hori and Kanaya's mortality solution is an interesting object of study, their application of this notion is misleading. I have no quarrel with their argument that if individuals take pleasure in doing the things they believe their dead ancestors would want them to do, then the eternity solution is appropriate even though the generations may not overlap. My disagreement is with their interpretation of the mortality solution. If I correctly interpret the passages quoted from Hori and Kanaya, they suggest that the mortality solution would be the appropriate utility function for individuals if generations did not overlap and if individuals believed that things that happen in the future can have no influence on the well-being of dead ancestors. In this case, I think that the two-sided altruism model is simply the wrong specification of interdependent utility. I suspect that the usual motivation for the two-sided altruism model is that the lives of parents and their children overlap and each generation takes pleasure in the other's happiness. If, as Hori and Kanaya suggest, generation $t - 1$ is already dead when generation t makes its choices and if the younger generation believes that the dead will never know the difference about what happens after they die, then it appears that the appropriate model of intergenerational benevolence is Barro's one-sided altruism model.

Although I don't believe that the Hori and Kanaya's mortality solution helps to answer the question: "How would the current generation evaluate allocations between itself and future generations if generations don't interlap and the dead are not conscious?," it does seem to me that this function helps to answer a different and quite interesting question. The mortality solution, which, as Hori and Kanaya observe, "regard(s) past generations' utility levels as givens," is an analytical device that can be used analogously to the way that demand theory uses "compensated demand curves." The mortality solution allows us to examine generation t's preferences among allocations that generation $t-1$ thinks are equally good. This can help us to answer the question: "Given that each generation *does* care about the well-being of its predecessor and of its successor as in the two-sided altruism model, what things do they agree about and what do they disagree about?" We see, for example, that consumption by generations earlier than t does not enter the mortality function given by equation (38) at all. This does not mean that the happiness of generation t is unaffected by the consumption of its predecessors, but rather that these effects are completely accounted for by the way in which they influence U_{t-1}. There is no disagreement between generations t and $t-1$ about which is the better of two allocations of consumption among generation $t-1$ and its predecessors *so long as the two allocations being compared do not differ in their treatment of generation t* and its successors. This should be no surprise, given the the result of Proposition 6 that in the two-sided altruism model there is backward-looking consistency. On the other hand, we also see from equation (38) that generation t is not *not* indifferent among *all* allocations that generation $t-1$ regards as indifferent. In particular, we see from this equation that if we define

$$D_t^+(c) = u(c_t) + \sum_{s=1}^{\infty} \beta^s u(c_{t+s}), \tag{41}$$

then among those allocations that generation $t-1$ views as indifferent, generation t prefers those with higher values of $D_t^+(c)$. The fact that the $D_t^+(c)$ is a positively weighted sum of subutilities of generations born at time t and later reflects the fact that generation t is more interested in the future and less interested in the past than generation $t-1$.

Though Hori and Kanaya's mortality solution expresses generation t's utility as a function of the utility of generation $t-1$ and of the consumptions of generation t and its successors, it is also possible to express the utility of generation t as a function of the utility of generation $t-1$ and of the consumptions of generation $t-1$ and its predecessors.[12]

In particular, we can show that if $a + b < 1$ then equation (31) describes the preferences of each generation t if and only if preferences of

[12]These results can be established with a proof that mirrors our proof of Proposition 7, exchanging the roles played by the matrices, $I - \alpha J^{-1}$ and $I - \beta J$ in the proof.

each individual t can also be represented by the bounded utility function

$$U_t = \beta U_{t+1} + u_t + \sum_{s=1}^{\infty} \alpha^s u(c_{t-s}), \tag{42}$$

where $\alpha = \alpha(a, b)$ and $\beta = \beta(a, b)$. Taking a lagged version of equation (42) in which t is replaced by $t - 1$ and rearranging terms, we find that

$$U_t = \frac{1}{\beta} U_{t-1} - \frac{1}{\beta} D_{t-1}^{-}(c), \tag{43}$$

where we define

$$D_{t-1}^{-}(c) = u(c_{t-1}) + \sum_{s=1}^{\infty} \alpha^s u(c_{t-1-s}). \tag{44}$$

Since equation (44) contains no terms involving the consumption of generations t and later, it must be that generations t and $t - 1$ are in complete agreement in their ranking of allocations that differ only in the consumptions of generation t and its successors. This is in accord with the result that in the two-sided altruism model there is forward-looking consistency of preferences between adjacent generations. We also see that for any two allocations that generation $t - 1$ thinks are equally good but that have different values of $D_{t-1}^{-}(c)$, generation t will prefer the one for which $D_{t-1}^{-}(c)$ is smaller.

Thus, we have two formulations of utility, both holding constant the utility of generation $t - 1$, and one shows that generation t prefers higher values of D_t^{+} and the other shows that generation t prefers lower values of $D_{t-1}^{-}(c)$. To see the connection between these results, notice that equations (33), (41), and (44) imply that

$$U_{t-1} = G_{t-1}(c) = D_{t-1}^{-}(c) + \beta D_t^{+}(c). \tag{45}$$

From Equation (45), we see that one can hold the utility of generation $t - 1$ constant while increasing $D_t^{+}(c)$ if and only if at the same time one decreases $D_{t-1}^{-}(c)$. Both equation (38) and equation (43) tell us that such a change, while leaving generation $t - 1$'s utility unaltered, will increase the utility of generation t.

6.3 Impatient Philanthropists

There are many possible extensions of the model of two-sided altruism. This is an example.

Suppose that each generation is concerned about the happiness of its predecessors and of its successors but is more concerned about consumption it can observe than about consumption it cannot observe. Although each generation likes its parent generation and its successor generation to be happy, it would prefer for them to obtain this happiness directly from their own consumption rather than indirectly from the happiness of more distantly removed generations. If this is the case, generations t and $t+1$ will in general disagree about transfers of consumption from generation $t+1$ to $t+2$, but will agree about changes in allocations that affect only generations $t+2$ or later.

This story would be consistent with the following system of interdependent utility functions for each generation t:

$$U_t = k_1 u_{t-1}(c_{t-1}) + u_t(c_t) + k_2 u_{t+1}(c_{t+1}) + aU_{t-1} + bU_{t+1}. \tag{46}$$

Equivalently, we can describe this system by the matrix equation

$$U = k_1 J^{-1} u(c) + u(c) + k_2 Ju(c) + AU, \tag{47}$$

where J is the matrix with 1's on the first superdiagonal and 0's elsewhere and J^{-1} is the matrix with 1's on the first subdiagonal and 0's elsewhere and where $A = aJ^{-1} + bJ$. Equation (47) implies that the vector utility function over allocations is given by the matrix equation

$$G(c) = (I - A)^{-1}(k_1 J^{-1} + I + k_2 J)\, u(c) \tag{48}$$

whenever $(I - A)^{-1}$ exists. In Proposition 5 we found $(I - A)^{-1}$ for the case where $a + b < 1$. From this solution we find that generation t's independent utility function for allocations is given by

$$G_t(c) = (k_1\alpha + 1 + k_2\beta)\, u_t(c_t) + \sum_{j=1}^{\infty} \alpha^j (k_1\alpha^{-1} + 1 + k_2\alpha)\, u_{t-j}(c_{t-j})$$
$$+ \sum_{j=1}^{\infty} \beta^j (k_1\beta + 1 + k_2\beta^{-1})\, u_{t+j}(c_{t+j}), \tag{49}$$

where, as previously, $\alpha = \alpha(a,b)$ and $\beta = \beta(a,b)$. One can divide this expression through by the positive number $k_1\alpha + 1 + k_2\beta$ to obtain an equivalent and more neatly expressed representation of the same preferences. This is

$$U_t = k_1^* \sum_{j=1}^{\infty} \alpha^j u(c_{t-j}) + u(c_t) + k_2^* \sum_{j=1}^{\infty} \beta^j u(c_{t+j}), \tag{50}$$

where

$$k_1^* = \frac{k_1\beta^{-1} + 1 + k_2\beta}{k_1\alpha + 1 + k_2\beta} \quad \text{and} \quad k_2^* = \frac{k_1\alpha + 1 + k_2\alpha^{-1}}{k_1\alpha + 1 + k_2\beta}. \tag{51}$$

Notice that with this system, the ratio of the weight that generation t puts on the consumption utilities of generation $t+1$ and $t+2$ is

$$\frac{k_2^*\beta}{\beta^2} = \frac{k_2^*}{\beta},$$

while the weight that generation $t+1$ puts on its own consumption utility relative to that of $t + 2$ is $1/\beta$. Therefore, generation t will disagree with generation $t + 1$ about how to allocate consumption between $t + 1$ and later generations. On the other hand, it is readily seen that generations t and $t + 1$ will be in agreement about allocations that differ only in the consumption bundles received by generations $t + 2$ or later.

7. Uniform Good Will

Let us consider a group of n persons such that every individual cares equally about the happiness of all other group members. The utility of individual i is given by

$$U_i = F_i(u_i(c_i), U_{\sim i}) = u_i(c_i) + a\sum_{j\neq i} U_j, \tag{52}$$

where $a > 0$. Where $u(c)$ is the column vector whose ith component is $u_i(c)$, this system of interdependent utility functions can be expressed by the matrix equation

$$MU = u(c), \tag{53}$$

where M is the $n \times n$ matrix with 1's on the diagonal and $-a$ in every off-diagonal entry. If M^{-1} exists, then the vector of independent utility functions over allocations is given by the matrix equation

$$G(c) = M^{-1}u(c). \tag{54}$$

It is easy to check that the matrix M is dominant diagonal if and only if $0 \leq (n-1)\,a < 1$, in which case M^{-1} is a positive multiple of a matrix with 1's on the diagonal and with each off-diagonal entry equal to $a/(1-(n-2)\,a)$. Then preferences of each individual i can be represented by the following independent utility function over allocations:

$$G_i(c) = u_i(c_i) + \alpha(a,n)\sum_{j\neq i} u_j(c_j), \tag{55}$$

where

$$\alpha = \alpha(a,n) = \frac{a}{1-(n-2)\,a}. \tag{56}$$

Since $\alpha(a, n)$ is an increasing function of n, we see that if the size of the group is increased and the coefficient a remains constant, then each individual would place a larger utility weight α on the private subutility of *each* other person relative to his own consumption.

An alternative hypothesis is that as group size increases, an individual's concern for the happiness of each member of the group is diluted.

For example, suppose that $a = k/(n-1)$ for some constant $k < 1$. Then the system of interdependent utilities can be written as

$$U_i = F_i(u_i(c_i), U_{\sim i}) = u_i(c_i) + k\bar{U}_{\sim i}, \tag{57}$$

where $\bar{U}_{\sim i}$ is the mean of U_j over all $j \neq i$. In this case,

$$\alpha\left(\frac{k}{n-1}, n\right) = \left(\frac{1}{n-1}\right)\left(\frac{k}{1 - k\left(\frac{n-2}{n-1}\right)}\right). \tag{58}$$

Where n is large, the expression in equation (58) is approximately

$$\left(\frac{1}{n-1}\right)\left(\frac{k}{1-k}\right).$$

Therefore, for large n the induced utility over allocations found in equation (55) is approximately

$$G_i(c) = u_i(c_i) + \frac{k}{1-k} - \bar{u}_{\sim i}(c), \tag{59}$$

where $\bar{u}_{\sim i}(c)$ is the mean of $u_j(c_j)$ over all $j \neq i$.

8. Conclusion

This paper examines systems of utility functions in which the utility functions of two or more people simultaneously depend on each others' utilities. In the cases studied, interdependent utility systems induce preferences of each individual over *allocations* of goods among the interrelated population. For the purposes of *demand theory*, it is the latter preferences that are of the greatest interest because it is simpler to study how one buys commodities for others than to study how one buys utility for others. On the other hand, introspection and casual observation often seem to offer stronger and more interesting hypotheses about the nature of interdependent utilities than they do about preferences over allocations. Hypotheses about the nature of utility interdependence imply special structure for the preference over allocations. The purpose of this paper has been to explain the way in which this happens.

Appendix

Proof of Lemma 1: Let A be a nonnegative denumerable matrix and let $(I - A)D$ have positive row sums for some bounded diagonal matrix D. This proof follows the proof for the finite-dimensional case, which was presented by Gale (1960). The only extra ingredient needed here that is not required for the finite-dimensional case is the boundedness of the matrix D. This is needed in order to show that $\lim_{n\to\infty}(I + A + A^2 + \ldots + A^n)$ exists.

If $I - A$ is dominant diagonal, then for some diagonal matrix D, and some $\mu > 0$, $(I - A)D > \mu I$. Let x^* be the vector whose ith component is the element D_{ii} of D. Then it must be that $Ax^* \ll \lambda x^*$ where $\lambda = 1 - \mu$. By induction it follows that $A^n x^* \ll \lambda^n x^*$.

For any integer n, define the matrix $B(n) = (I + A + A^2 + \ldots + A^n)$. Since the matrix A is nonnegative it must be that for all i and j, $B_{ij}(n+1) \geq B_{ij}(n)$. Furthermore, because $Ax^* \ll \lambda x^*$, it follows that for all n,

$$B(n)x^* \leq (1 + \lambda + \lambda^2 + \ldots + \lambda^n)x^* < \frac{1}{1-\lambda}x^*.$$

But because $x^* \gg 0$ is a bounded vector and because $B(n) \geq 0$ for all n, it follows that the sequence $B_{ij}(n)$ is bounded. Since a bounded monotonic sequence must converge to a limit, it follows than $\lim_{n\to\infty} B_{ij}(n) = B_{ij}$ exists and hence the matrix $B = \lim_{n\to\infty} B(n)$ is well defined. It is straightforward to verify that $B \geq 0$ and $B(I - A) = (I - A)B = I$.

We also need to show that B is the only bounded inverse for $(I - A)$. If B and B' are both bounded inverses, then $B(I - A) = B'(I - A) = (I - A)B = (I - A)B' = I$. Then $(I - A)(B - B') = 0$. Since B and B' are, by assumption, bounded matrices, the matrix $B - B'$ also is bounded. Therefore every column of $B - B'$ is a bounded vector. Suppose that $(I - A)x = 0$ for a bounded vector, x. Where $B(n) = (I + A + A^2 + \ldots + A^n)$, $0 = B(n)(I - A)x = (I - A^n)x$. From our argument of the previous paragraph it follows that $\lim_{n\to\infty} A^n = 0$. Therefore, if x is a bounded vector it must be that $x = 0$. From this it follows that every column of $B - B'$ is a zero vector and hence that $B = B'$.

The converse result that if $(I - A)$ has a nonnegative bounded inverse, B, then $I - A$ must be dominant diagonal can be shown as follows. Let e be the column vector with 1's in every element. Then let $y = Be$ be a bounded vector. Since all entries of B are nonnegative, it must be that $y \geq 0$. Since B is an inverse of $I - A$, it must be that $(I - A)y = e \gg 0$. From Remark 1 it follows that $I - A$ is dominant diagonal.[13] ■

Proof of Lemma 2: We first show that F maps $\mathcal{X}$ into $\mathcal{Y}$. Since $a + b < 1$ for $(a, b) \in \mathcal{X}$, it must be that $ab < a(1 - a)$ and, hence, that

$$0 \leq (1 - 2a)^2 = 1 - 4a(1 - a) < 1 - 4ab.$$

[13] I am grateful to H. Hori for pointing out an error in my original proof of this result.

Therefore it must be that $1 - 2a < \sqrt{1-4ab}$ and, hence, that

$$0 \leq \beta(a,b) = \frac{1-\sqrt{1-4ab}}{2a} < 1.$$

A symmetric argument shows that for $(a,b) \in \mathcal{X}$ it must also be that $0 \leq \alpha(a,b) < 1$. Thus $F(a,b) \in \mathcal{Y}$ for all $(a,b) \in \mathcal{X}$.

Next we show that for any $(\alpha,\beta) \in Y$, there is one and only one (a,b) in X such that $(\alpha,\beta) = F(a,b)$. Calculation shows that if $\alpha = \alpha(a,b)$ and $\beta = \beta(a,b)$, then

$$1 + \alpha\,\beta = \frac{1-\sqrt{1-4ab}}{2ab},$$

and furthermore that $a = \alpha/(1+\alpha\beta)$ and $b = \beta/(1+\alpha\beta)$.

Therefore the function F has a well-defined inverse and

$$F^{-1}(\alpha,\beta) = \left(\frac{\alpha}{1+\alpha\beta}, \frac{\beta}{1+\alpha\beta}\right). \tag{60}$$

Finally, we show that for for all (α,β) in $\mathcal{Y}$, $F^{-1}(\alpha,\beta)$ is in the set $\mathcal{X}$. From equation 60 we see that where $(a,b) = F^{-1}(\alpha,\beta)$, it must be that $a \geq 0,\ b \geq 0$, and

$$a + b = \frac{\alpha+\beta}{1+\alpha\beta}. \tag{61}$$

Subtracting the numerator from the denominator of the right-hand side of equation 61, we find that $1 + \alpha\beta - (\alpha+\beta) = (1-\alpha)(1-\beta) \geq 0$. Therefore it must be that $a + b \leq 1$ and, hence, that $(a,b) \in \mathcal{X}$.

It remains to be shown that F is continuous. Obviously F is continuous at (a,b) where $a > 0$ and $b > 0$. All that we need to show is that for if $0 \leq b < 1$, $\lim_{b\to 0}\alpha(a,b) \to a$, and $\lim_{a\to 0}\beta(a,b) \to b$. This follows from direct application of L'Hospital's rule. ■

Proof of Lemma 3: The first assertion is immediate from the definitions. The second assertion can be proved by direct computation. For $a > 0$, $b > 0$, and $a + b < 1$, it must be that $0 < \alpha(a,b) < 1$ and $0 < \beta(a,b) < 1$. Calculating the partial derivatives of $\alpha(a,b)$ with respect to its first and second arguments and rearranging terms, we find that

$$\alpha_1(a,b) = \frac{1}{\sqrt{1-\alpha(a,b)\beta(a,b)}} > 0 \tag{62}$$

and

$$\alpha_2(a,b) = \frac{\beta(a,b)^2}{\sqrt{1-\alpha(a,b)\beta(a,b)}} > 0. \tag{63}$$

A symmetric argument shows that the partial derivatives $\beta_1(a,b)$ and $\beta_2(a,b)$ are also positive. ■

Proof of Lemma 4: Multiplying the terms on the right-hand side of equation (35), we have

$$I + \frac{\alpha}{1+\alpha\beta} J^{-1} + \frac{\beta}{1+\alpha\beta} J = I + aJ^{-1} + bJ. \tag{64}$$

■

Proof of Proposition 5: From Lemma 2 it follows that $\beta < 1$ and $\alpha < 1$ and hence $(I - \beta J)$ and $(I - \alpha J^{-1})$ are dominant diagonal matrices.

The inverses of these matrices are seen to be

$$(I - \alpha J)^{-1} = I + \sum_{t=1}^{\infty} \alpha^t (J^{-1})^t \quad \text{and} \quad (I - \beta J)^{-1} = I + \sum_{t=1}^{\infty} \beta^t J^t. \tag{65}$$

From equations (35) and (65), it follows that

$$(I - A)^{-1} = (1 + \alpha\beta)(I - \beta J)^{-1}(I - \alpha J^{-1})^{-1} \tag{66}$$

$$= (1 + \alpha\beta)\left(I + \sum_{t=1}^{\infty} \beta^t (J^{-1})^t\right)\left(I + \sum_{t=1}^{\infty} \beta^t J^t\right). \tag{67}$$

Multiplying the matrices that appear on the right-hand side of equation (67), we find that

$$(I - A)^{-1} = \frac{1 + \alpha\beta}{1 - \alpha\beta} B, \tag{68}$$

where B is the matrix such that $B_{ij} = \beta^{j-i}$ for $j \geq i$ and $B_{ij} = \alpha^{i-j}$ for $i \geq j$.[14] Thus the independent utility functions induced by two-sided altruism are represented by the vector equation

$$U = \frac{1 + \alpha\beta}{1 - \alpha\beta} Bu. \tag{69}$$

[14]There are also some *unbounded* matrices B such that $U = Bu$ is a solution to $(I - A)U = u$. To see where these come from, recall that when we parameterized the system with $a = \alpha/(1 + \alpha\beta)$ and $b = \beta/(1 + \alpha\beta)$, we found that there was exactly one solution for (α, β) such that $0 \leq \alpha < 1$ and $0 \leq \beta < 1$. But there is also a solution, (α', β') where $0 \leq \alpha' < 1$ and $0 \leq \beta' < 1$. The matrix $(I - aJ^{-1} - bJ)$ can also be factored into the expression $(I - \alpha' J^{-1})(I - \beta' J)$. But when the inverse of this matrix is computed, it turns out to be unbounded.

Therefore, preferences of any each individual t can be represented by the utility function

$$U_t = \left(\frac{1+\alpha\beta}{1-\alpha\beta}\right)\left(\sum_{j=1}^{\infty} \alpha^j u_{t-j}(c_{t-j}) + u_t(c_t) + \sum_{j=1}^{\infty} \beta^j u_{t+j}(c_{t+j})\right). \quad (70)$$

Since $0 < \alpha < 1$ and $0 < \beta < 1$, it must be that $[(1+\alpha\beta)/(1-\alpha\beta)] > 0$. Therefore the utility function in equation (33) of Proposition 5 is a monotonic transformation of that in equation (70) and hence represents the same preferences. ■

Proof of Proposition 6: If we fix consumption of each each generation $i \leq t$ at $\bar{c}_i$ and consider allocations that differ only in c_s for $s > t$, we see from Proposition 5 that preferences of generation t can be represented by a utility function of the form

$$\tilde{U}_t(c_{t+1}, \ldots, c_{t+s}, \ldots) = C + \sum_{s=1}^{\infty} \beta^s u\,(c_{t+s}), \quad (71)$$

where $C = u(c_t) + \sum_{s=1}^{\infty} \alpha^{-s} u(c_{t-s})$ and preferences of generation $t+1$ among such bundles can be represented by a utility function of the form

$$\tilde{U}_{t+1}(c_{t+1}, \ldots, c_{t+s}, \ldots) = C' + \sum_{s=1}^{\infty} \beta^{s-1} u(c_{t+s}), \quad (72)$$

where $C' = \alpha u(c_t) + \sum_{s=2}^{\infty} \alpha^{-s} u(c_{t-s})$. From their definitions, we see that

$$\tilde{U}_{t+1}(c_{t+1}, \ldots, c_{t+s}, \ldots) = C' - \beta C + \beta \tilde{U}_t(c_{t+1}, \ldots, c_{t+s} \ldots). \quad (73)$$

Therefore the utility function (72) is an increasing (affine) transformation of the utility function (71) and hence the two utility functions must represent the same preferences. This establishes forward-looking consistency between adjacent generations.

A symmetric argument proves backward-looking consistency between adjacent generations. ■

Proof of Proposition 7: From Lemma 4 it is immediate that $(I - A)U = u$ if and only if

$$(I - \alpha J^{-1})\,U = (1+\alpha\beta)(I - \beta J)^{-1} u. \quad (74)$$

If $0 \leq \beta < 1$, then it is easily seen that $(I - \beta J)^{-1}$ exists and is equal to the matrix $I + \sum_{s=1}^{\infty} \beta^s J^s$ which has 0's below the diagonal, 1's on the diagonal, and entries β^s on the sth superdiagonal. Therefore it must be that

$$U = \alpha J^{-1} + (1+\alpha\beta)\left(I + \sum_{s=1}^{\infty} \beta^s J^s\right) u. \quad (75)$$

Calculating the tth row from equation (75), we find that

$$U_t = \alpha U_{t-1} + u_t + \sum_{s=1}^{\infty} \beta^s u_{t+s}. \tag{76}$$

It remains to be demonstrated that $0 \le \beta < 1$ if and only if either $a + b < 1$ or if $a + b > 1$ and $b < 1/2$. For the independent utility functions found in equation (5) to be bounded, it was necessary both that $\alpha(a, b) < 1$ and that $\beta(a, b) < 1$. Lemma 2 shows that $0 \le \alpha(a, b) < 1$ and $0 \le \beta(a, b) < 1$ if and only if $a + b < 1$. The current theorem requires only that $0 \le \beta(a, b) < 1$ and this admits some additional solutions. In particular, the equation

$$I - aJ^{-1} - bJ = (1 + \alpha\beta)(I - \alpha J^{-1})(I - \beta J) \tag{77}$$

has two solutions:

$$\alpha = \frac{1 - \sqrt{1 - 4ab}}{2b} \quad \text{and} \quad \beta = \frac{1 - \sqrt{1 - 4ab}}{2b} \tag{78}$$

or

$$\alpha = \frac{1 + \sqrt{1 - 4ab}}{2b} \quad \text{and} \quad \beta = \frac{1 + \sqrt{1 - 4ab}}{2b} \tag{79}$$

Calculation shows that where α and β solve equations (40), it must be that $\alpha\beta > 1$, which implies that either $0 \le \alpha > 1$ and $0 \le \beta > 1$. There are, however, some values of a and b for which $\alpha > 1$ and $0 \le \beta < 1$. In particular, $0 \le \beta < 1$ if and only if $a + b > 1$ and $b < 1/2$. ■

References

BARRO, R. (1974) Are government bonds net wealth?, *Journal of Political Economy* **82**, 1095–1117.

BERGSTROM, T. (1989) Love and spaghetti: The opportunity cost of virtue, *Journal of Economic Perspectives* **3**, 165–173.

BLACKORBY, C., D. NISSEN, D. PRIMONT, and R. RUSSELL (1973) Consistent intertemporal decision making, *Review of Economic Studies* **40**, 239–248.

BUITER, W., and J. CARMICHAEL (1984) Government debt: Comment, *American Economic Review* **74**, 762–765.

BURBIDGE, J. (1983) Government debt in an overlapping generations model with gifts and bequests, *American Economic Review* **73**, 222–227.

BURBIDGE, J. (1984) Government debt: Reply, *American Economic Review* **75**, 766–767.

DEBREU, G. (1960) Topological methods in cardinal utility theory; in *Mathematical Methods in the Social Sciences*, K. ARROW, S. KARLIN, and P. SUPPES, eds. Stanford, CA, Stanford University Press, 16–26.

GALE, D. (1960) *The Theory of Linear Economic Models.* New York: McGraw-Hill.

HORI, H. (1989) Utility functionals with nonpaternalistic intergenerational altruism: The case where altruism extends to many generations. *Journal of Economic Theory* **56**, 451–467.

HORI, H., and S. KANAYA (1989) Utility functions with nonpaternalistic intergenerational altruism. *Journal of Economic Theory* **49**, 241–255.

KEMENY, J., J. L. SNELL, and A. KNAPP (1966) *Denumerable Markov Chains.* Princeton, N.J.: Van Nostrand.

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

KOOPMANS, T. C. (1960) Stationary ordinal utility and impatience, *Econometrica* **28**, 287–309.

KOOPMANS, T. (1972) Representation of preference orderings over time; in *Decision and Organization, A Volume in Honor of Jacob Marschak*, C. McGUIRE and R. RADNER, eds. Amsterdam: North Holland, 79–100.

McKENZIE, L. (1960) Matrices with dominant diagonal and economic theory; in *Mathematical Methods in the Social Sciences*, K. ARROW, S. KARLIN, and P. SUPPES, eds. Stanford, CA: Stanford University Press, 47–62.

NIKAIDO, H. (1968) *Convex Structures and Economic Theory.* New York: Academic Press.

STROTZ, R. (1955) Myopia and inconsistency in dynamic utility maximization, *Review of Economic Studies* **23**, 165–180.