JOURNAL OF ECONOMIC THEORY **61**, 189–205 (1993)

Intertemporal Transfer Institutions

JOAN MARIA ESTEBAN AND JÓZSEF SÁKOVICS*

Institut d'Anàlisi Econòmica (CSIC), Campus Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

Received March 12, 1992; revised October 1, 1992

We consider a single commodity overlapping generations model where, at a cost, institutions can be created (or transformed) to carry out intergenerational transfers. We analyze the game with both strategic and cooperative methods, characterizing the "unique" stationary equilibrium, as well as the set of transfer institutions that belong to the core (which we show to coincide with the consistent core). We conclude that, as long as the creation of an institution is costly, it is possible to sustain positive transfers, though those will be below the optimal level. Moreover, the lower the costs, the more efficient the transfer will be. *Journal of Economic Literature* Classification Numbers: C71, C72, D23, E21, H55. © 1993 Academic Press, Inc.

1. INTRODUCTION

In this paper we examine the social viability of intergenerational transfers in a model of Overlapping Generations (OG). Since the first OG model was laid down by Samuelson [16], it is known that competitive equilibria may be dynamically inefficient and would never entail a positive amount of intergenerational transfer. Further, it is also well known that the introduction of valueless fiat money may bring about the existence of Pareto efficient monetary equilibria with a positive price for money. Intergenerational transfers are thus possible when appropriately backed with some implicit social contract. Samuelson himself called this the "social contrivance of money." But, as pointed out in his seminal paper, "the Golden Rule of Kant's Categorical Imperative (enjoining like people to follow the common pattern that makes each best off) is often not self-enforcing." As a matter of fact, it was Hume [9] who first made the point that human society is composed of overlapping generations of citizens and that the newly born individuals could not feel obliged by a pre-existing social contract they did not even implicitly agree upon.[1]

* We are grateful to Carmen Matutes for her suggestions. Financial support has been provided by the Spanish Ministry of Education through Grant DGICYT PB90-0172.

[1] Hume's argument was later taken up by Bentham [1].

0022-0531/93 $5.00

The social viability of intergenerational implicit contracts has been recently analyzed from both the cooperative and noncooperative points of view. Hendricks *et al.* [8] and Esteban [3] have proven that in a one-commodity OG pure exchange economy, no monetary equilibrium belongs to the core. Later, Esteban and Millan [4] showed that this negative result is robust: despite the fact that with many commodities and many agents some monetary equilibria can belong to the core, as the economy becomes large all those equilibria are eventually excluded from the set of core allocations.

The noncooperative case was first examined in a one-commodity OG model by Hammond [6] and later by Kandori [10] and Salant [15]. It is easy to see that when the set of strategies is not history-dependent and consists of either performing the prescribed transfer or defecting, the dominant strategy equilibrium is defection. However, Hammond in his "pension game," as well as Kandori and Salant, show that, using strategies depending on the full history, with punishment rules for deviant generations it is possible to sustain all individually rational intergenerational transfers as a subgame-perfect equilibrium. We find that the recourse to players with unbounded memory is not very insightful in this specific environment. Thus, our endeavor is to present a model where transfers are self-enforced (arise and are maintained naturally) instead of being enforced by the subsequent generations.[2]

It is our view that societies have solved the need for intergenerational trust by creating social artifacts called "institutions." As North [13] has recently put it, "institutions are the humanly devised constraints that structure political, economic and social interaction. They consist of both informal constraints (sanctions, taboos, customs, traditions, and codes of conduct), and formal rules (constitutions, laws, property rights)." The role of institutions is to crystallize agreements by means of built-in rules that make eventual changes costly. Political constitutions are a good example of the social device we have in mind. They have a built-in mechanism that makes it quite cumbersome to change the rights and duties embodied in them. The autonomy of monetary authorities with respect to the government also illustrates the point.

In this paper we examine the role of institutions in supporting intergenerational transfers.[3] We use the canonical one-commodity, pure exchange, stationary overlapping generations model with agents who live

[2] To that effect we restrict our attention to Markov strategies, where, conditional on the previous transfer, the agents' choice of action is independent of the past. (See also the discussion following Definition 1.)

[3] The role of institutions, laws, and social rules in making viable intergenerational transfers has already been examined by Engineer [2], Veall [17], Kotlikoff *et al.*, [12], and Hansson and Stuart [7], among others.

for two periods. Institutions are designed to carry out transfer policies for the entire sequence of generations. A particular transfer policy is a single number, independent of the history, that represents the transfer from young to old.[4] Individual generations can amend the existing institution and define a new one, which they themselves are also supposed to abide by, performing an alternative intertemporal sequence of transfers. Moreover, we assume that changing the existing institution has a fixed cost.

We study the emergence and viability of institutions from both the non-cooperative and the cooperative approaches and obtain remarkably similar results. The availability of institutions (with non-zero change costs) makes intergenerational transfers viable. Specifically, we obtain that the golden rule transfer always belongs to the core (which we show to coincide with the consistent core) and that the set of core transfers coincides with the maximal set of fixed points of the equilibria in stationary strategies of the non-cooperative game. The results from the cooperative and non-cooperative approaches agree in predicting that the core, and thus the set of fixed points of the stationary equilibria, shrinks to the golden rule transfer as the fixed cost becomes small. However, the two approaches diverge in that in the limit the core contains the golden rule only, while under strategic behavior we obtain many other equilibria, including the no-transfer institution. Finally, our results predict that if the economy starts with no intergenerational transfer, it will eventually reach a stationary positive transfer, though an inefficient one, below the golden rule level.

The paper is organized as follows. After outlining the model in the next section, Section 3 is devoted to the case of noncooperative behavior, while in Section 4 we examine the core of the game. The conclusions are drawn in the last section.

2. The Model

We assume an infinite horizon economy in which time is divided into discrete periods of constant length. In every period $t \geqslant 0$ a new generation of agents is born and lives for two periods, t and $t+1$. Generations are composed of one single agent each. Thus, at any point in time the generation just born (T) will coexist with the old generation ($T-1$). In period $t=0$ there also exists a generation of old agents, born in

[4] In principle, the transfer policy could be more complex but, in addition to the technical and expositional complications that it would create, we believe that the appropriate model of a transfer policy in a stationary environment is the constant one.

the previous (unmodeled) period. We assume that there is one single commodity, which cannot be stored from one period to the next. We denote by $a(t)$ and $b(t+1)$ the consumption of this commodity by agent T in his youth and in his old days, respectively. Individual preferences will be assumed to be identical over generations and representable by the utility function[5]

$$U(a(t), b(t+1)) = a(t)\, b(t+1).$$

Each individual receives endowments $\omega > 1$ and 1 in the first and second period of his life, repectively. Therefore, the environment is stationary. The commodity is transferable between coexisting generations. To simplify matters we assume that only the young can transfer to the old, in line with the assumption that the young have the larger endowment. Intergenerational transfers can be performed through *institutions*. Institutions can be set up, transformed, or terminated at a cost. We identify institutions by the transfer they perform, $x \in [0, \omega]$. Whenever the institution is altered (that is, the size of the transfer is changed), the corresponding (giving) generation will have to bear the cost of that change. This cost could possibly depend on the size of the institution which is to be dismantled and of the new one that is being set up, $c(x_{t-1}, x_t)$. In this paper we assume, however, that $c(x_{t-1}, x_t) = c \geqslant 0$ whenever $x_{t-1} \neq x_t$, and $c(x_{t-1}, x_t) = 0$ otherwise.[6] An institutional transfer x is *feasible* if $0 \leqslant x \leqslant \omega$ and it is *constrained feasible* if $0 \leqslant x \leqslant \omega - c \equiv \hat{\omega}$. We denote these sets by F and CF, respectively. We assume that the set of constrained feasible transfers, CF, is non-empty, that is, $c < \omega$. We use as benchmarks two institutional transfers: the *golden rule* and the *constrained golden rule*, x^* and $\hat{x}$, respectively. These transfers are defined by

$$(\omega - x^*)(1 + x^*) \geqslant (\omega - x)(1 + x) \qquad \forall x \in F \tag{1}$$

and

$$(\hat{\omega} - \hat{x})(1 + \hat{x}) \geqslant (\hat{\omega} - x)(1 + x) \qquad \forall x \in CF. \tag{2}$$

It is routine work to obtain that $x^* = (\omega - 1)/2$ and $\hat{x} = (\hat{\omega} - 1)/2$. Further, it is immediate that since $c \geqslant 0$, we have that $\hat{x} \leqslant x^*$.

The *Pareto efficient* consumption allocations will correspond to the ones generated by transfers x satisfying $x \geqslant x^*$. It is essential to our model that

[5] Our results can be easily extended to the more general class of preferences satisfying concavity (in the transfer) and that the marginal rate of substitution along the 45 degree line is unity. In our view the slim gain in generality does not compensate for the increase in clumsiness.

[6] Kotlikoff *et al.* [12] also assume a fixed cost for rewriting a social contract.

intergenerational transfers can only be carried out by means of costly institutions. Therefore, it is natural to introduce the notion of constrained efficiency, taking into account the actual feasibility constraints. In that case, the *constrained Pareto efficient* consumption allocations will be generated by transfers satisfying $x \geqslant \hat{x}$.

3. Noncooperative Analysis

We start by the strategic analysis of our model. In this section, we assume that agents cannot sign enforceable contracts or engage in any form of explicit cooperation. They single-mindedly try to maximize their own utility, taking into account, however, the effects of their actions on other players' behavior, which has thus an indirect effect on their own payoff.

Definition 1. *We define a (pure) strategy of player T to be a function that maps the level of the transfer institution inherited from player $T-1$ to the new transfer institution. A mixed strategy is then defined as*

$$V_T\colon [0, \omega] \to \mathscr{F},$$

where $\mathscr{F}$ is the set of probability distributions over $[0, \omega]$.

Note that these are Markov-type strategies, since an agent's action depends only on the current institution and not on the full history of the game. We make this restriction for several reasons: First, we are using non-cooperative methods partly to find out what happens if people are only concerned with what happens to them and not even indirectly with what happens to others. Second, given the time horizon for any application, the assumption that players remember the full history, and condition their actions on it, is quite unrealistic. Third, we want to liberate ourselves from the Folk-theorem-like results of repeated games, where by using "punishment strategies" an embarrassingly large number of equilibria can be supported, rendering the predictions useless. Finally, we are interested in modeling intergenerational transfers via institutions. In this context it is a natural assumption that agents will condition their behavior on the existing institution only.[7]

As the solution concept we use the notion of subgame-perfect (Nash) equilibrium. Lemma 1 demonstrates that the difficulties that could arise from the fact that in the model there are an infinite number of agents moving sequentially can be circumvented, and there exists a simple recursive characterization of equilibrium.

[7] Consider, if you will, the current institution as a "sufficient statistic" of the full history.

LEMMA 1. *An infinite sequence of strategies,* $V_i(\cdot)$ $i = 0, 1, 2, \ldots,$ *constitutes a subgame-perfect (Nash) equilibrium if and only if the following holds for every* $t(=T)$ *and every* $x_{t-1} \in [0, \omega]$:

$$\operatorname{supp} V_T(x_{t-1}) \subseteq \arg\max_{x_t} \, (\omega - c(x_{t-1}, x_t) - x_t)(1 + EV_{T+1}(x_t))$$

$$s.t. \; \omega - c(x_{t-1}, x_t) - x_t \geqslant 0.$$

Proof. Note that the above condition states that T's strategy has to be a best response to $T+1$'s strategy (subject to the feasibility constraint). Since T's payoff is independent of the actions of all subsequent players, such a strategy is a best response to the strategies of agents $T+1$, $T+2$, ... On the other hand, since the previous agents' actions are summarized by the current institution, the optimal action of the current agent will by definition be a best response to the previous players' strategy. Subgame perfection follows from the fact that each agent moves only once and that we check optimality in exactly that subgame. Q.E.D.

At this level of generality we can say little about the set of equilibrium strategies or that of equilibrium paths. Therefore, in what follows we restrict our attention to equilibria where the agents' strategies are stationary, or, in other words, they are the same:[8] $V_T(y) \equiv V_{T+1}(y)$ $\forall T \geqslant 0$. Given the fact that all agents have identical preferences and are in the same situation (that is, the environment is stationary), and our previous assumption that they use Markov-type strategies, so their actions are not dependent on the full history, it seems intuitive to require them to act independently of their place in the stream of generations.

Let $\delta(x) = 0$ or x or 1.[9] Also, define $\alpha_y = (y - c)/y(\omega - y)$ and $A = \sqrt{(x^* + 1)^2 - \hat{\omega}}$.

PROPOSITION 1. *If* $\hat{\omega} \leqslant 1$ *then the "unique" stationary equilibrium is the following:*[10]

$$V(y) = \begin{cases} y, & \text{if } 0 \leqslant y < c \\ \delta(\alpha_y) \otimes y \oplus (1 - \delta(\alpha_y)) \otimes 0, & \text{if } c \leqslant y \leqslant x^* + A \\ 0, & \text{if } x^* + A < y \leqslant \omega. \end{cases}$$

The proof is in the Appendix.

[8] Note that this assumption does not require the equilibrium sequence of transfers to be stationary, it only refers to the decision rules.

[9] That is, in each equilibrium it has to take exactly one out of the three possible values.

[10] We denote a mixed strategy that puts weight α on action x and weight $1 - \alpha$ on action y by $\alpha \otimes x \oplus (1 - \alpha) \otimes y$.

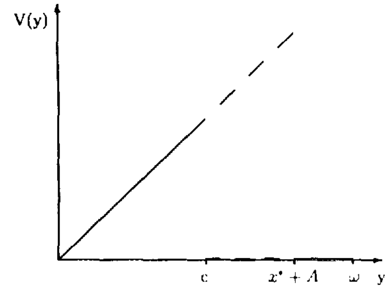

FIG. 1. The stationary equilibrium strategy if $\hat{\omega} \leqslant 1$.

Thus, if the cost of changing the institution is more than the difference between the endowments the agents are getting in their two periods of life ($c \geqslant \omega - 1$), then the only feasible change is to dismantle the institution completely, since after paying the cost of change the remaining endowment in the first period is less than the second period endowment and therefore an intergenerational transfer from young to old is not efficient. Consequently, if the existing transfer is less than c it is never optimal to change. If the current transfer is very large then the savings achieved by not carrying out the transfer compensate for the cost of change and the fact that the agent will not receive a transfer in his old days. There is also a transitionary region[11] of the existing transfers, where both strategies are optimal, and there is even an equilibrium with in strategies.

Let

$$\beta_y = \frac{(\hat{x}+1)(\hat{x}+1-\omega+y)}{(\omega-y)(y-\hat{x})}.$$

Also let

$$B = \sqrt{(x^*+1)^2-(\hat{x}+1)^2}.$$

PROPOSITION 2. *If* $1 < \hat{\omega} < \omega$ *then the "unique" stationary equilibrium is the following*:

$$V(y) = \begin{cases} \beta_y \otimes y \oplus (1-\beta_y) \otimes \hat{x}, & \textit{if} \quad 0 \leqslant y < x^* - B \\ y, & \textit{if} \quad 0 \leqslant x^* - B \leqslant y < \hat{x} + c \\ \delta(\beta_y) \otimes y \oplus (1-\delta(\beta_y)) \otimes \hat{x}, & \textit{if} \quad \hat{x} + c \leqslant y \leqslant x^* + B \\ \hat{x}, & \textit{if} \quad x^* + B < y \leqslant \omega. \end{cases}$$

[11] Note that $c < x^* + A$ for $c \in (0, \omega)$.

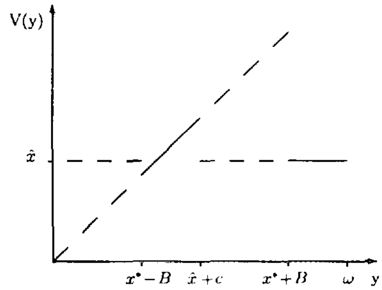

FIG. 2. The stationary equilibrium strategy if $1 < \hat{\omega}$.

The proof is in the Appendix.

Thus, when the cost of changing the institution is sufficiently small then, if the transfer is changed at all, it is changed to the constrained golden rule, $\hat{x}$. If the existing transfer is close to $\hat{x}$ then it is not worthwhile to change. If the existing transfer is low[12] then a mixed strategy must[13] be used, since if the agent's strategy were to change to $\hat{x}$ then the previous generation would prefer not to change, and, similarly, if the current agent was planning to leave the institution intact, the previous one would rather change it.

We can also see in this case that, if we start out with zero initial transfer and $c < (\sqrt{\omega} - 1)^2$ then with probability 1 eventually the institution will be built, to a level of $\hat{x}$, at which level it will thereafter be maintained. Since $\hat{x} < x^*$ this means that even in the steady state the transfer carried out will fall below the socially optimal level. It is also worth noting, however, that if some external force initiated the economy with a transfer institution that carries out the golden rule transfer, in equilibrium no generation would want to change the institution, regardless of the size of $c > 0$.

Finally, if change is costless, we lose uniqueness:

COROLLARY 1 (to Lemma 1). *If the change of institutions is costless* ($c = 0$) *and* $V(z)$ *is a stationary equilibrium, then*

$$
\begin{gathered}
x \in \operatorname{supp} V(z) \\
\Updownarrow \\
x \in \arg\max(\omega - x)(1 + EV(x)).
\end{gathered}
\tag{3}
$$

The corollary directly follows from Lemma 1. The difficulties arising from the lack of an institution (an institution that is costless to change has no

[12] As well as the cost, for note that if $c > (\sqrt{\omega} - 1)^2$ then $x^* - B < 0$.

[13] Note that in the other region where a mixed strategy may be used, there are also two pure strategy equilibria. Here the mixed strategy is unique.

effect on the agents' behavior) are twofold: first, it is easy to see from the corollary that there exists a large number of stationary equilibria; second, the "mistrust strategy" ($V(y) \equiv 0$) is a stationary equilibrium.

4. Cooperative Behavior

We now examine the outcome of cooperative behavior. In this framework, agents can lead an objection to the ruling institutional transfer by initiating a new size of transfer and reneging from the old one. In that case, that leading generation will bear the cost of substituting one institution for another. Agents can object to the existing transfer when there is another institution such that, taking into account both its cost and the transfers it performs, it makes both the leader and all subsequent generations strictly better off. Then, we shall say that the core is the set of un-objected institutional transfers.

Since we are assuming a stationary environment and stationary institutions, it is clear that if an institution is objected to by a coalition headed by generation T, it will also be objected to by the infinite sequence of coalitions we can form by choosing $T = 0, 1, 2, \ldots$ as the leading generation. We can therefore concentrate on coalitions headed by generation $T = 0$.

Let us proceed to a formal definition of blocking by means of coalitions and of the core as the set of unblocked institutions.

Definition 2. *We say that coalition S blocks institution x with y if:*

(i) $(\omega - c - y)(1 + y) > (\omega - x)(1 + x)$, *for the leading generation, and*

(ii) $(\omega - y)(1 + y) > (\omega - x)(1 + x)$, *for all the subsequent generations.*

The core is the set C of unblocked institutional transfers.

For $c \geqslant 0$ it is immediate that $(\omega - y)(1 + y) \geqslant (\omega - c - y)(1 + y)$. Therefore, condition (i) in Definition 2 becomes the only binding inequality.

Our notion of the core deserves some comments. When looking at cooperative outcomes, we are interested in those allocations at which no subset of agents finds it advantageous to exclude the rest of agents while reallocating the available resources within the coalition. This notion can be made rigorous in a number of ways, according to the object of study and the assumed environment. In our case, individuals are supposed to choose among alternative institutions available to them. Therefore the object of cooperation or exclusion are the institutions themselves. Here we are identifying institutions with the transfers they are designed to carry out.

Therefore, when a generation decides to break away and set up a new institution, that generation reneges from the transfer commitment implicit in the previous institution, pays the transformation costs, and starts abiding by the intergenerational transfer rules associated with the new institution just created.

Let us start by characterizing the core.

PROPOSITION 3. *The institutional transfer x belongs to the core, $x \in C$, if and only if $x \in [x_m, x_M]$, where*[14]

$$x_m = \max(0, x^* - B), \tag{4}$$

and

$$x_M = \begin{cases} x^* + A, & \text{if } c \geqslant \omega - 1 \\ x^* + B, & \text{if } c < \omega - 1. \end{cases} \tag{5}$$

Proof. Let us start by proving sufficiency: It is easy to check that $x \in [x_m, x_M]$ implies that $(\omega - x)(1 + x) \geqslant \max_{z \geqslant 0}(\hat{\omega} - z)(1 + z)$. Therefore such an institution x cannot be blocked, so it belongs to the core.

To verify necessity, consider first the case when $c \geqslant \omega - 1$. In this case $x_m = 0$, so we only need to show that no $x > x^* + A$ belongs to the core. But for such an x, $(\omega - x)(1 + x) < \hat{\omega}$ and therefore the no-transfer institution would block any such transfer. If $(\sqrt{\omega} - 1)^2 \leqslant c < \omega - 1$ then x_m is still zero, so we need to show that no $x > x^* + B$ is in C. But for such an x $(\omega - x)(1 + x) < (\hat{\omega} - \hat{x})(1 + \hat{x})$ and therefore the constrained golden rule transfer, $\hat{x}$, is blocking. Finally, if $0 \leqslant c < (\sqrt{\omega} - 1)^2$ then x_M is the same as before, while $x_m = x^* - B$. Thus we only need to show that no $x < x^* - B$ belongs to the core. But for such an x $(\omega - x)(1 + x) < (\hat{\omega} - \hat{x})(1 + \hat{x})$ and therefore the constrained golden rule transfer is blocking. Q.E.D.

We now turn to existence.

PROPOSITION 4. *C is nonempty.*

Proof. From the definitions of x_m and x_M, (4) and (5), it is trivial to observe that $x^* \in [x_m, x_M]$ for all $0 \leqslant c \leqslant \omega$ and hence $x^* \in C$.

COROLLARY 2. *If $c = 0$ then C is a singleton.*

Proof. For $c = 0$, $\hat{x} = x^* = x_m = x_M$.

The bounds x_m and x_M, and thus the size of the core set, depend on the value of c. The lower bound, x_m, is decreasing and the upper bound, x_M,

[14] Recall that $A = \sqrt{(1 + x^*)^2 - \hat{\omega}}$ and $B = \sqrt{(1 + x^*)^2 - (1 + \hat{x})^2}$.

is increasing in c. It is easy to compute that for all c such that $(\sqrt{\omega}-1)^2 \leqslant c < \omega$ the no-transfer institution $x=0$ belongs to the core. In order to object to an existing institution, the leading generation of the blocking coalition has to choose an alternative institution from the set of CF institutions. By Corollary 2 when $c=0$, $CF=F$ and only the golden rule (x^*) belongs to the core. On the other hand, as c increases, CF shrinks and as $c \to \omega$ it converges to a singleton set containing $x=0$ only. In that case, the set of feasible objections reduces to the no-transfer institution and correspondingly the core coincides with the set of feasible transfers, F.

Heretofore, we have described the economy from an arbitrary starting time, with a pre-existing institution. It seems natural to address the question of how this institution came about, or, in other words, whether all institutions are equally likely to be observed at a random time t. This is equivalent to examining what would be the choice of a generation when faced with an existing institution not belonging to the core. Note that a relevant special case is when the economy starts with $x=0$, and this does not belong to the core. From the previous analysis it is clear that $\max(0, \hat{x})$ is the most preferred institution in CF. But since $\max(0, \hat{x}) < x^*$ for $c > 0$, the most likely institution to be observed will carry out an inefficient (too low) intergenerational transfer. Therefore, either an external decision maker has chosen at some point in the past a Pareto efficient institution that belongs to the core, or else the blocking generation will always choose a core, but Pareto inefficient, transfer except for the exceptional case $c=0$.

Let us now examine whether the core as defined here strictly refines the set of solutions satisfying the different consistency conditions that have recently been explored in the literature. The approach initiated by Ray [14] and followed by Esteban and Ray [5] is to introduce the notion of credibility, which requires for a blocking objection raised by coalition S_r involving all generations, $t \geqslant r$, to be credible in the sense than it cannot be blocked by any other subcoalition S_s, $s > r$, raising an objection itself credible. With a countable infinity of agents, this approach presents an obvious problem of infinite regress. In Esteban and Ray [5] this is solved by defining k-credible objections and the corresponding k-consistent core, $C(k)$, and letting k tend to infinity. Let us introduce the formal definition. Bear in mind, though, that because of the infinite horizon, the set of actions open to any coalition S_t is identical to the set available to the full sequence of players as seen at $t=0$.

Definition 3. *We say that an institutional transfer x is k-credibly blocked by S_r with the transfer y, if $y \in C(k-1)$ and condition* (i) *of Definition 2 is satisfied. The transfer x belongs to the k-consistent core if it cannot be k-credibly blocked by any coalition. Whenever*

$\limsup_{k \to \infty} C(k) = \liminf_{k \to \infty} C(k)$ *the Consistent Core,* (CC) *exists and is equal to this limit.*

Khan and Mookherjee [11] follow a different route, much closer in spirit to the von Neumann–Morgenstern solution set. We paraphrase their definition below. Let us consider the set of feasible institutional transfers F.

DEFINITION 4. *We say that the sets G and B are a consistent partition of F if*

(a) $x \in B \Leftrightarrow \exists y \in G$ *such that* y *blocks* x,

(b) $x \in G \Leftrightarrow \nexists y \in G$ *such that* y *blocks* x.

If the set F can be consistently partitioned then G is the set of Consistent Coalition Proof Equilibria $(CCPE)$.

It is clear from this definition that here we are requiring the blocking allocations to belong to a set in which allocations cannot block each other.

PROPOSITION 5. $C = CC = CCPE$.

Proof. In Proposition 3 we have fully characterized the core, that is, the set $C(0)$. Let us now examine the set $C(1)$, which obtains by restricting the potential blocking coalitions to object with some $y \in C(0)$. We have now that every $x \notin C(0)$ can be blocked with $\hat{x} \in C(0)$. Therefore, $C(1) = C(0)$. It follows that $C(k) = C(k-1)$ for all $k = 1, 2, \ldots$ and therefore that CC exists and $CC = C$.

Let us now check whether the core coincides with $CCPE$. We start by noting that the sets C and $\bar{C}$ form a partition. Let us first check B. Assume that $x \in \bar{C}$. Then we know from Proposition 3 that x is blocked by $\hat{x} \in C$. Let us now suppose that x is blocked by some $y \in C$. Then x cannot belong to C, that is, $x \in \bar{C}$. Hence $x \in B \Leftrightarrow x \in \bar{C}$.

Let us now check for G. Assume that $x \in C$. Then by the definition of C there is no $y \in C$ blocking x. Consider now that there is no $y \in C$ blocking x. Then $x \in C$, since by Proposition 3 all $x \in \bar{C}$ are blocked by $\hat{x} \in C$. Therefore, $x \in G \Leftrightarrow x \in C$. Hence $CCPE = C$. Q.E.D.

In our model institutions provide an intergenerational link in two ways. On the one hand, when agents set up a new institution they start by performing the transfer in benefit of the current old generation whose institution has been rejected. On the other, we have the inertial role played by the existence of costs to institutional change, which may restrain agents from blocking the ruling institution. The former is responsible for the core being non-empty, while the latter for its size. Indeed, when the leading generation does not have to start by performing the new transfer the core is empty, as shown in Hendricks *et al.* [8] and Esteban [3]. Not even the golden rule transfer is in the core because each generation finds it advantageous to lead a blocking coalition proposing the same golden rule scheme but delayed for one period. In this way they expect to receive a transfer without having made one.

5. Conclusion

We have presented a model of intergenerational transfer institutions that gives an intuitive argument for the existence of intergenerational transfers (for example, pensions), explains their relative size, and also justifies the need for institutions to carry them out. We have isolated a single characteristic of institutions as the significant factor, their inertia. By building an institution today we can ensure that future generations be committed to it, at least to the degree of the costliness of implementing a change. This implicit commitment is sufficient to steer us away from the disheartening result of mistrust: no transfer at all. Once we believe (with reason) that the next generation will make us a transfer if we do, it becomes optimal to perform. In fact, we have shown even more: if the setup costs are not prohibitive, it is optimal to perform a transfer (that is, to create an institution) even if the previous generation did not "oblige" us to make a transfer. Thus, we have an explanation for the emergence of institutions: the efficiency loss from the lack of enforceable cooperation is so important that it pays to incur a significant cost in order to ensure at least a form of implicit cooperation.

We have made full use of the fact that our setup is amenable to both strategic and cooperative analysis. The results obtained by the two approaches are strikingly(?) in line with one another. For non-zero costs of change the transfers belonging to the (consistent) core are exactly[15] the fixed points of the stationary equilibrium. That is, once a level of transfer that belongs to the core is established, both methods of analysis predict that the institution will be kept intact. By Corollary 2 this interval always[16] includes the golden rule transfer, conforming with the intuition that the most efficient transfer should not be changed. In general, the set of stationary transfers is an interval around x^*, the golden rule, corresponding to the fact that if we are "close enough" to the optimum, then it does not pay to incur the fix cost of change. As the cost reaches the level of the first period endowment, all transfers become stationary: people cannot afford to change.

[15] Assuming that a generation will not change the institution if it is indifferent.

[16] For positive costs, in the non-cooperative case.

The two approaches only differ in the benchmark case of zero cost. While the cooperative outcome predicts the golden rule transfer as the unique outcome (since the golden rule is the socially optimal outcome, any coalition prefers it to any other institutional program), in the non-cooperative case we get multiple equilibria, including the dreaded no-transfer equilibrium. In this instance we side with the conclusion of the strategic analysis. When costs are positive then the cooperative method distinguishes the leading generation of the blocking coalition by the fact that it will have to absorb the full cost of change. However, when the cost is zero, the coalition is symmetric, the pivotal role of the leading generation gets lost in contrast to our understanding of the situation. On the other hand, the noncooperative method and result agree both with our intuition and existing results in the literature. Therefore, we conclude that costly institutions are necessary to ensure positive intergenerational transfers.

What about the evolution of institutions? If costs are not too high $(0 < c < (\sqrt{\omega} - 1)^2)$, then 0 is not a stationary transfer. The cooperative approach predicts a change to one of the core transfers, while the strategic analysis predicts a (eventual) change to the constrained golden rule. In this case the non-cooperative method refines the cooperative result. Since the constrained golden rule is strictly less then the golden rule, this means that the transfers we are likely to observe are strictly below optimal. We can also conclude that the smaller the inertia of an institution, the closer to the efficient level its transfer will be. Therefore, if a social planner were to decide on the preferred cost of change then, since as we have seen, the zero cost case is not desirable, she would find that, though the smaller the cost the better, there does not exist an (socially) optimal cost level. That is, costly transfer institutions are useful but they should be as flexible as possible.

Appendix

The Proof of Proposition 1

Optimality. Let us assume agent $T+1$'s strategy is as specified in the proposition and let us verify that it is a best reponse to itself:

If $y < c$ then T has two options: either he leaves the institution unchanged and earns $(\omega - y)(1 + y)$ or he changes the transferred amount. In this second case the optimal new transfer is $\arg\max_{x \geqslant 0}(\hat{\omega} - x)(1 + EV(x)) = \arg\max_{x \geqslant 0}(\hat{\omega} - x)(1 + x) = 0$, since $\hat{\omega} \leqslant 1$. Thus it is optimal not to change the transfer if $(\omega - y)(1 + y) \geqslant \hat{\omega}$. But this is implied by $y < c$.

If $c \leqslant y \leqslant x^* + A$ then T's options are again either to leave the institution intact and to earn $(\omega - y)(1 + \delta(\alpha_y) y)$ or to change it to zero and to earn $\hat{\omega}$. But note that if $\delta(\alpha_y) = 0$ then $(\omega - y)(1 + \delta(\alpha_y) y) = \omega - y \leqslant \hat{\omega}$, so abolishing the institution is optimal; if $\delta(\alpha_y)=1$ then $(\omega-y)(1+y)>\hat{\omega}$, so leaving it intact is a best response, while if $\delta(\alpha_y)=\alpha_y$ then $(\omega-y)(1+\alpha_y y)=\hat{\omega}$, for all y, therefore mixing between 0 and y is optimal.

If $y>x^*+A$ then T's options are again either to leave the institution intact and to earn $\omega-y$ or to change it to zero and to earn $\hat{\omega}$. Now, if $y>c$ then $\omega-y\leqslant\hat{\omega}$, so it is optimal to abolish the institution.

Uniqueness. It is straightforward to verify that if $V(\cdot)$ is a stationary equilibrium then it has to satisfy the following for all $y\in[0,\omega]$:

$$\begin{aligned}\operatorname{supp} V(y)\subseteq\{y\mid(\omega-y)(1+EV(y))\geqslant\sup_{x\geqslant 0}(\hat{\omega}-x)(1+EV(x))\}\\ \cup\{\arg\max_{0\leqslant x\neq y}(\hat{\omega}-x)(1+EV(x))\mid(\omega-y)(1+EV(y))\\ \leqslant(\hat{\omega}-x)(1+EV(x))\}.\end{aligned}\tag{6}$$

That is, if the existing institution is y then in order for a transfer to be in the support of a stationary equilibrium strategy, either it has to be the same transfer y yielding at least as much utility as any possible other transfer or it has to be the best possible new transfer that on top gives a higher payoff than sticking to the old institution.

Let $\bar{x}\in\arg\max_{x\geqslant 0}(\hat{\omega}-x)(1+EV(x))$. (If the maximum does not exist then by (6) $\operatorname{supp} V(y)\subseteq\{y\}$.) Then $(\omega-\bar{x})(1+EV(\bar{x}))>\max(\hat{\omega}-x)(1+EV(x))$ and therefore $V(\bar{x})=\bar{x}$. But then

$$(\hat{\omega}-\bar{x})(1+EV(\bar{x}))=(\hat{\omega}-\bar{x})(1+\bar{x}).\tag{7}$$

If $\hat{\omega}\leqslant 1$ then (7) has its unique maximum at $\bar{x}=0$. Thus the second set in (6) is either empty or it contains 0 only. Consequently, $\operatorname{supp} V(y)\subseteq\{0,y\}$. Then by (6), $0\in\operatorname{supp} V(y)$ only if $(\omega-y)(1+y)\leqslant\hat{\omega}$ and similarly $y\in\operatorname{supp} V(y)$ only if $\omega-y\geqslant\hat{\omega}$. In the transitionary region it is easy to verify that there are three possible solutions: $V(y)=0$, $V(y)=y$, and $V(y)=\alpha_y\otimes y\oplus(1-\alpha_y)\otimes 0$. Q.E.D.

The Proof of Proposition 2

Optimality. Let us assume agent $T+1$'s strategy is as specified in the proposition and let us verify that it is a best response to itself:

If $y<x^*-B$ then T has two options: either he leaves the institution unchanged and earns (in expectation) $(\omega-y)(1+\beta_y y+(1-\beta_y)\hat{x})$ or he changes the transferred amount. In this second case the optimal new transfer is

$$\arg\max\Big(\max_{s<x^*-B}X(s),\ \max_{x^*-B\leqslant s<\hat{x}+c}Y(s),\ \max_{\hat{x}+c\leqslant s\leqslant x^*+B}Z(s)\Big),$$

where $X(s)=(\hat{\omega}-s)(1+\beta_s s+(1-\beta_s)\hat{x})$, $Y(s)=(\hat{\omega}-s)(1+s)$, and $Z(s)=$

$(\hat{\omega}-s)(\delta(\beta_s)s+(1-\delta(\beta_s))\hat{x})$. It is clearly suboptimal to increase the transfer above x^*+B since changing it to $\hat{x}$ would give the same payoff at a lower cost.

Now, observe that $(\omega-s)(1+\beta_s s+(1-\beta_s)\hat{x})=(\hat{\omega}-\hat{x})(1+\hat{x})=\max Y(z)$ for all $s<x^*-B$. Therefore, $\max_{s<x^*-B} X(s)<\max Y(z)$ since $\hat{\omega}<\omega$. Similarly, it is easy to see that $\max Y(z)>Z(s)$ for all $s>\hat{x}+c$. Moreover, T will be indifferent between his two best choices (keeping the old institution or changing the transfer to $\hat{x}$) since both give him $(\hat{\omega}-\hat{x})(1+\hat{x})$. Consequently, it is optimal to randomize.

If $x^*-B\leqslant y<\hat{x}+c$ then T has two options: either he leaves the institution unchanged and earns $(\omega-y)(1+y)$ or he changes the transferred amount. In this second case the optimal new transfer is $\hat{x}$ as before. Then, it is optimal not to change the transfer (to $\hat{x}$) if $(\omega-y)(1+y)\geqslant(\hat{\omega}-\hat{x})(1+\hat{x})$. But this is implied by $y<\hat{x}+c<x^*+B$.

If $\hat{x}+c\leqslant y\leqslant x^*+B$ then T has two options: either he leaves the institution unchanged and earns $(\omega-y)(1+\delta(\beta_y)y+(1-\delta(\beta_y))\hat{x})$ or he changes the transferred amount. In this second case the optimal new transfer is again $\hat{x}$. Thus if $\delta(\beta_y)=0$ it is optimal to change the transfer to $\hat{x}$ since then $(\omega-y)(1+\delta(\beta_y)y+(1-\delta(\beta_y))\hat{x})\leqslant(\hat{\omega}-\hat{x})(1+\hat{x})$. If $\delta(\beta_y)=1$ then similarly it is optimal not to change the institution. Finally, if $\delta(\beta_y)=\beta_y$ then T will be indifferent between his two options.

If $y>x^*+B$ then T's options are again either to leave the institution intact and to earn $(\omega-y)(1+\hat{x})$ or to change it to $\hat{x}$ and to earn $(\hat{\omega}-\hat{x})(1+\hat{x})$. But when $y>x^*+B$, $(\omega-y)(1+\hat{x})<(\omega-y)(1+y)<(\hat{\omega}-\hat{x})(1+\hat{x})$, so it is optimal to change the transfer to $\hat{x}$.

Uniqueness. If $1<\hat{\omega}\leqslant 2\sqrt{\omega}-1$ then (7) has its unique maximum at $\bar{x}=\hat{x}$. Thus the second set in (6) is either empty or contains $\hat{x}$ only. Consequently, $\operatorname{supp} V(y)\subseteq\{\hat{x}, y\}$. Then by (6), $\hat{x}\in\operatorname{supp} V(y)$ only if $(\omega-y)(1+\max(\hat{x}, y))\leqslant(\hat{\omega}-\hat{x})(1+\hat{x})$ and similarly $y\in\operatorname{supp} V(y)$ only if $(\omega-y)(1+\min(\hat{x}, y))\geqslant(\hat{\omega}-\hat{x})(1+\hat{x})$. Thus, if $y<\hat{x}+c$ then $V(y)=y$ while if $y\geqslant x^*+B$ then $V(y)=\hat{x}$. Finally, in the transitionary region it is easy to verify that there are three possible solutions: $V(y)=\hat{x}$, $V(y)=y$, and $V(y)=\beta_y\otimes y\oplus(1-\beta_y)\otimes\hat{x}$.

If $\hat{\omega}>2\sqrt{\omega}-1$ then (7) still has its maximum at $\bar{x}=\hat{x}$ and thus $\operatorname{supp} V(y)=\{\hat{x}, y\}$. The conditions for $y\in\operatorname{supp} V(y)$ and $\hat{x}\in\operatorname{supp} V(y)$ are also the same: $y\in\operatorname{supp} V(y)$ only if $(\omega-y)(1+\max(\hat{x}, y))\leqslant(\hat{\omega}-\hat{x})(1+\hat{x})$ while $\hat{x}\in\operatorname{supp} V(y)$ only if $(\omega-y)(1+\min(\hat{x}, y))\geqslant(\hat{\omega}-\hat{x})(1+\hat{x})$. The difference is that now the equation $(\omega-y)(1+y)=(\hat{\omega}-\hat{x})(1+\hat{x})$ has two nonnegative solutions $y=x^*\pm B$. Thus for $y\geqslant x^*-B<\hat{x}+c$ $V(y)$ must be the same as before. If $y<x^*-B$ then it is easy to verify that only the strictly mixed strategy giving weight β_y to y is a stationary equilibrium. Q.E.D.

References

1. J. Bentham, "A Fragment on Government," 1776.
2. M. Engineer, "Essays on the Structure of Financial Institutions," Ph.D. thesis, Queens University, 1987.
3. J. Esteban, A characterization of the core in Overlapping Generations economies, *J. Econ. Theory* **39** (1986), 439–456.
4. J. Esteban and T. Millan, Competitive equilibria and the core of Overlapping Generations economies, *J. Econ. Theory* **50** (1990), 155–174.
5. J. Esteban and D. Ray, "Credible Coalitions and the Consistent Core in Overlapping Generations Economies," Working paper, IAE-Universitat Autonoma de Barcelona, 1991.
6. P. Hammond, Charity: Altruism or cooperative egoism? *in* "Altruism, Morality and Economic Theory" (E. S. Phelps, Ed.), Russell Sage Foundation, New York, 1975.
7. I. Hansson and C. Stuart, Social security as a trade among living generations, *Amer. Econ. Rev.* **79** (1989), 1182–1195.
8. K. Hendricks, K. Judd, and D. Kovenock, A note on the core of the Overlapping Generations model, *Econ. Lett.* **6** (1980), 95–97.
9. D. Hume, "Essays Moral, Political and Literary," 1758.
10. M. Kandori, Repeated games played by Overlapping Generations of players, *Rev. Econ. Stud.* **59** (1992), 81–92.
11. C. M. Khan and D. Mookherjee, The Good, the Bad and the Ugly: Coalition proof equilibrium in infinite games, *Games Econ. Behav.* **4** (1991), 101–121.
12. L. Kotlikoff, T. Persson, and L. Svensson, Social contracts as assets: A possible solution to the time-consistency problem, *Amer. Econ. Rev.* **78** (1988), 662–677.
13. D. C. North, Institutions, *J. Econ. Persp.* **5** (1991), 97–112.
14. D. Ray, Credible coalitions and the core, *Int. J. Game Theory* **18** (1989), 185–187.
15. D. J. Salant, A repeated game with finitely lived Overlapping Generations of players, *Games Econ. Behav.* **3** (1991), 244–259.
16. P. A. Samuelson, An exact consumption-loan model of interest with or without the social contrivance of money, *J. Polit. Econ.* **66** (1958), 467–482.
17. M. R. Veall, Public pensions as optimal social contracts, *J. Public Econ.* **31** (1986), 237–251.