Maastricht University

P. Jean-Jacques Herings, Arkadi Predtetchinski

Procedurally Fair Taxation

RM/11/024

METEOR

Maastricht University School of Business and Economics
Maastricht Research School of Economics
of Technology and Organization

P.O. Box 616
NL - 6200 MD Maastricht
The Netherlands

Procedurally Fair Taxation

P. Jean-Jacques Herings* Arkadi Predtetchinski†

March 29, 2011

Abstract

We study the implications of procedural fairness on income taxation. We formulate procedural fairness as a particular non-cooperative bargaining game and examine the stationary subgame perfect equilibria of the game. The equilibrium outcome is called tax equilibrium and is shown to be unique. The procedurally fair tax rate is defined as the tax rate that results in the limit of tax equilibria when the probability that negotiations break down converges to zero. The procedurally fair tax rate is shown to be unique. We also provide a characterization of the procedurally fair tax rate that involves the probability mass of below average income citizens and a particular measure of the citizens' boldness. This characterization is then used to show that in a number of interesting cases the procedurally fair tax rate equals the probability mass of below average income citizens.

KEYWORDS: Procedural fairness, income taxation, bargaining.

JEL CODES: C78, D63, H24.

*P.J.J. Herings, Department of Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands. This author would like to thank the Netherlands Organization for Scientific Research (NWO) for financial support. P.Herings@maastrichtuniversity.nl

†A. Predtetchinski, Department of Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands. This author would like to thank the Netherlands Organization for Scientific Research (NWO) for financial support. A.Predtetchinski@maastrichtuniversity.nl

1 Introduction

Modern societies redistribute significant shares of their total income to improve upon the fairness of the income distribution. At the same time, the literature in economics offers very few guidelines to determine what fraction of total income should be redistributed if the end result has to be fair. The standard approach follows Mirrlees (1971) and consists of specifying a social welfare function that has to be optimized, without explaining where the social welfare function comes from and how the social welfare function should be chosen. A notable exception is Fleurbaey and Maniquet (2006), who derive the social welfare function on the basis of the Pigou-Dalton transfer principle and a condition precluding redistibution when all agents have the same skills.

One may distinguish fair allocations from fair procedures to decide upon the allocation. Bolton, Brandts, and Ockenfels (2005) argue that procedural fairness is conceptually distinct from allocation fairness, although the two are linked in important ways. This paper is concerned with procedural fairness. Procedures are deemed fair if they create equal chances for persons involved in the procedure. It is equal opportunities that matters rather than an equal allocation that results. Fair procedures are often used in decision making when there is no clear candidate for a fair allocation. An example of a fair procedure is where an allocation of goods is based on a random lottery. In this case, using a fair procedure comes at a huge cost, the loss of efficiency.

In this paper we consider a fair procedure, called the unanimous approval procedure, that does not suffer from efficiency losses. In the unanimous approval procedure, all citizens in the society have an equal chance of making a proposal. A proposal is only carried out if it is approved by everyone, which is in accordance with the contractarian approach. If one of the citizens rejects, each citizen has the same chance to become the new proposer, and so on, and so forth. The unanimous approval procedure is a natural extension of the alternating offers bargaining model of Rubinstein (1982) to the case with more than two players. A difference is that in the unanimous approval procedure the citizens do not alternate or rotate in making proposals, but are selected with equal probability in every bargaining round, thereby following the approach of Binmore (1987). The unanimous approval procedure follows as a special case of the bargaining models of collective choice considered in Banks and Duggan (2000), when we restrict recognition probabilities to be uniform and replace approval by a coalition within a collection of decisive coalitions by unanimous approval. Contrary to the approach in Banks and Duggan (2000), we intend to use bargaining models of collective choice as normative tools here.

We apply the unanimous approval procedure to income taxation. We consider a society with citizens that are characterized by their productivity level and their utility function, where the distribution of characteristics is given by some probability measure. Due to

the heterogeneity in productivity and preferences, there is no clear candidate for a fair allocation. Neither is there a clear candidate for a social welfare function.

We analyze the case where citizens use the unanimous approval procedure to decide upon the desired tax rate. Each citizen has an equal chance to make a proposal, a proposal of a citizen specifies a tax rate, and a particular proposal is only implemented if all citizens approve of it. One may think of this as a stylized way to model an ideal direct democracy or as a stylized way to establish a social contract. We assume tax rates to be non-negative. For reasons of incentive compatibility, after-tax income should be increasing in pre-tax income, which implies that tax rates are less than or equal to one. A tax rate of zero corresponds to the situation of laissez-faire and a tax rate of one to complete redistribution.

Citizens with below average pre-tax income would prefer complete redistribution, whereas citizens with above average pre-tax income have a tax rate of zero as their most preferred tax policy. Since tax rates are chosen in the unit interval, our analysis leads to the analysis of one-dimensional bargaining problems. In the bargaining and political economy literature, such problems are also studied in Banks and Duggan (2000), Imai and Salonen (2000), Cho and Duggan (2003), Kalandrakis (2006), Cardona and Ponsatí (2007), Predtetchinski (2007), and Herings and Predtetchinski (2010), though none of these models captures the specification presented here.

After each round of the unanimous approval procedure, there is some probability that negotiations break down. We define a tax equilibrium as the outcomes induced by stationary subgame perfect equilibria of the unanimous approval procedure for a given breakdown probability. A tax equilibrium is characterized by the unique proposal made by all above average income citizens and the unique proposal made by all below average income citizens. It is shown that in equilibrium the proposals made are accepted unanimously without delay and that tax equilibria are unique. Below average income citizens propose a strictly higher tax rate than above average income citizens.

Next we define the procedurally fair tax rate as the limit of the proposals in a tax equilibrium when the breakdown probability converges to zero. We show that every society has a uniquely determined procedurally fair tax rate. The procedurally fair tax rate admits a characterization in terms of the probability mass of citizens with below average income and an appropriate measure of boldness of the citizens. As shown in Roth (1989), boldness at a certain consumption level corresponds to the maximum probability by which a citizen prefers a particular gamble over getting the consumption level for sure. Boldness equals the first derivative of the utility function divided by utility itself. Tax rate boldness is defined as boldness applied to the indirect utility function in terms of tax rates. We argue that the bargaining power of the below average income citizens at a particular tax rate is equal to the supremum of tax rate boldness among them multiplied by their probability mass. The

bargaining power of the above average income citizens is defined similarly. We demonstrate that the procedurally fair tax rate is the unique tax rate for which the bargaining power of the below average income citizens is equal to that of the above average income citizens.

Our characterization of the procedurally fair tax rate is extremely helpful in computations. For instance, in societies where productivity levels have unbounded support and all citizens have the same preferences exhibiting constant relative risk aversion, the procedurally fair tax rate is given by the probability mass of below average income citizens.

We also consider heterogeneous societies, where the only assumption on preferences is that both among below average income citizens and among above average income citizens, there are citizens that are close to being risk neutral. We obtain a simple expression for the procedurally fair tax rate, irrespective of the distribution of productivity levels. When this distribution has unbounded support, we find again that the procedurally fair tax rate is given by the probability mass of below average income citizens.

The remainder of this paper is organized as follows. Section 2 introduces the primitives of our society and Section 3 discusses the unanimous approval procedure. Section 4 defines the notion of tax equilibrium, studies the relation to stationary subgame perfect equilibria of the unanimous approval procedure when applied to our society, and derives some properties of tax equilibria. Section 5 defines the procedurally fair tax rate and provides a characterization, which is used in Section 6 to argue that in sufficiently heterogeneous societies the procedurally fair tax rate is equal to the probability mass of below average income citizens. Section 7 concludes.

2 The Society

We consider a society composed of citizens in a set N, which may contain finitely or infinitely many elements. The characteristics of citizens are described by $(w_t, u_t)_{t\in N}$ and a probability measure μ. A citizen of type $t \in N$ has productivity $w_t \in \mathbb{R}$. He derives no disutility from labor, works one unit of time, and has a pre-tax income of w_t. It would be natural to assume that w_t is non-negative, but in order to simplify some of the proofs in this paper, it is helpful to allow for the more general case where pre–tax income could be negative.

The triple $(N, \mathcal{A}, \mu)$ denotes a probability space, where N is the set of citizens, $\mathcal{A}$ a sigma–algebra of subsets of N, and μ a probability measure that represents the distribution of types within the population. We assume w_t to be measurable and integrable and we let

$$\bar{w} = \int_{t\in N} w_t d\mu(t).$$

denote the average income. Furthermore, we denote the types with below average, average,

and above average income by N^-, N^0, and N^+, respectively. We make the regularity assumption that the probability is zero that a citizen has productivity exactly equal to the average level, $\mu(N^0) = 0$.

ASSUMPTION 2.1: The function w_t is measurable and integrable. Moreover, $\mu(N^0) = 0$, $\mu(N^-) > 0$, and $\mu(N^+) > 0$.

We let $m^- = \mu(N^-)$ and $m^+ = \mu(N^+)$ be the probability mass of citizens with below and above average income, respectively.

We address the question of the share of total income that should be redistributed in order to reach a procedurally fair income distribution. To do so we consider all possible affine tax functions, characterized by parameters α and β. After-tax income $\alpha + (1-\beta)w_t$ of a citizen of type t is an affine function of pre-tax income, and β is equal to the tax rate. Under the requirement of a balanced budget, we find that

$$\alpha + (1-\beta)\bar{w} = \bar{w},$$

so $\alpha = \beta\bar{w}$. The tax rate β coincides with the share of total pre-tax income that is redistributed. As a function of the tax rate β, after-tax income is equal to

$$\beta\bar{w} + (1-\beta)w_t = w_t + \beta(\bar{w} - w_t).$$

Not all tax rates β are feasible. We assume that β is greater than or equal to 0, a requirement that would follow if incomes are required to be non-negative and the infimum of productivity levels in the support of μ is equal to zero. For reasons of incentive compatibility, after-tax income should be increasing in pre-tax income, which implies that β is less than or equal to one.

When $\beta = 0$ we obtain the situation of laissez-faire, where after-tax income is equal to pre-tax income. The case $\beta = 1$ corresponds to complete redistribution. Since we will assume utility functions to be strictly increasing, a citizen's preferred point is laissez-faire when his income is above $\bar{w}$ and complete redistribution for income below $\bar{w}$. A household with average income is indifferent with regard to the tax policy chosen.

A citizen of type t evaluates the after–tax income using a von Neumann–Morgenstern utility function $u_t : [\ell_t, +\infty) \to \mathbb{R}$. Since we have allowed for negative pre–tax income, we also allow for a negative ℓ_t rather than simply imposing $\ell_t = 0$.

ASSUMPTION 2.2: For each $t \in N$ the utility function u_t is continuous, concave, strictly increasing, and $u_t(\ell_t) = 0$. Moreover, $\ell_t \leq w_t$ if $t \in N^-$, $\ell_t \leq \bar{w}$ if $t \in N^+$, and $\ell_t < \bar{w}$ if $t \in N^0$.

If the tax rate $\beta \in [0,1]$ is agreed upon, citizen t enjoys a utility of

$$v_t(\beta) = u_t(w_t + \beta(\bar{w} - w_t)).$$

The function $v_t : [0,1] \to \mathbb{R}$ defined by the above equation is the *indirect utility function* of type t. The indirect utility function v_t is non–negative, continuous, and concave. Moreover, it holds that $v_t(\beta) > 0$ whenever $0 < \beta < 1$. For each $t \in N^0$, v_t is a positive constant. For each $t \in N^-$, v_t is strictly increasing while for each $t \in N^+$ it is strictly decreasing.

Consider some $t \in N^- \cup N^+$. We denote the inverse of the utility function u_t by x_t. For υ a feasible utility level of citizen t, we define

$$h_t(\upsilon) = \frac{w_t - x_t(\upsilon)}{w_t - \bar{w}}.$$

The function h_t, restricted to the domain of feasible indirect utility levels, is the inverse of the indirect utility function v_t. For each $t \in N^-$ the function h_t is strictly increasing and convex while for each $t \in N^+$ it is strictly decreasing and concave.

3 The Unanimous Approval Procedure

The citizens in the society face a set of feasible tax rates given by $B = [0,1]$. To determine the procedurally fair tax rate, we study what tax rate will be chosen when citizens use the *unanimous approval procedure.* The unanimous approval procedure is defined as follows. In every period, each citizen has an equal chance to be selected as the proposer. More precisely, we have that a citizen of type t is selected as the proposer according to the probability measure μ. If in period τ a citizen of type t is selected as the proposer, he makes a proposal $p_t \in B$. After observing p_t, citizens sequentially decide whether to accept or to reject the proposal in a fixed a priorily chosen order. The latter assumption is made for notational simplicity and can be generalized substantially without affecting our results. If all citizens accept, then p_t is implemented and the utility of a citizen of type i is given by $v_i(p_t)$. As soon as a citizen rejects, the procedure breaks down with probability $1 - \delta$ and all citizens receive a breakdown utility of 0. With probability δ the procedure is repeated in period $\tau + 1$ and starts with the selection of a proposer. We are interested in the proposals that are made in the limit when δ converges to one.

The unanimous approval procedure leads to a well-defined game in extensive form. We analyze the subgame perfect equilibria in stationary strategies of this game. A *stationary strategy* of a citizen of type t, $\sigma_t = (p_t, A_t)$, consists of a *proposal* $p_t \in B$ and an *acceptance set* $A_t \subset B$. The acceptance set A_t consists of those proposals that are accepted by a citizen of type t. This specification results in a stationary strategy because p_t and A_t are time and

history independent. The *social acceptance set* consists of the proposals that are accepted by all citizens and is given by $A = \cap_{t\in N} A_t$.

Under appropriate measurability conditions, a strategy profile $\sigma = (\sigma_t)_{t\in N}$ determines a unique probability measure over the tax rates that are implemented and thereby the expected utility $U_t(\sigma)$ for each citizen of type t as evaluated at the beginning of the game.[1] Since strategies are stationary, $U_t(\sigma)$ is also the continuation utility of a citizen of type t, the expected utility as evaluated at the beginning of any time period τ.

A strategy profile σ is a *subgame perfect equilibrium in stationary strategies* (SSPE) if in any subgame the strategy profile σ induces a Nash equilibrium, i.e. for all $t \in N$, p_t is optimal given the strategies of all citizens, and A_t is optimal given the strategies of all citizens. It is standard in the literature to restrict attention to SSPE with the property that a citizen who is indifferent between accepting and rejecting a proposal will accept, and we call such an equilibrium SSPEI. We will define the *procedurally fair tax rates* as those that are proposed in an SSPEI when the continuation probability δ tends to one.

4 Tax Equilibrium

In this section we consider a strategy profile $\sigma = (p_t, A_t)_{t\in N}$ that is SSPEI and provide its characterization, called *tax equilibrium.* We show that tax equilibria exist and are unique.

Let $N^{\mathrm{a}} = \{t \in N \mid p_t \in A\}$ denote the set of citizens whose proposal is accepted. Citizen i's expected utility $U_i(\sigma)$ of the strategy profile σ satisfies the equation

$$U_i(\sigma) = \int_{t\in N^{\mathrm{a}}} v_i(p_t)d\mu(t) + (1-\mu(N^{\mathrm{a}}))\delta U_i(\sigma), \quad i \in N. \tag{4.1}$$

Recalling our assumption that a citizen who is indifferent between accepting and rejecting a proposal will accept, it follows that citizen i accepts the proposal β if and only if the utility $v_i(\beta)$ he derives from β is at least as large as the continuation utility. Hence

$$A_i = \{\beta \in [0,1] : v_i(\beta) \geq \delta U_i(\sigma)\}, \quad i \in N. \tag{4.2}$$

THEOREM 4.1: *Consider a strategy profile $\sigma = (p_t, A_t)_{t\in N}$ and the induced social acceptance set A. If σ is an SSPEI then*

[A] *The strategy profile σ has a no–delay property: $p_i \in A$ for every $i \in N$.*

[B] *The interval $[\delta\bar{p}, 1-\delta+\delta\bar{p}]$ is contained in A, where $\bar{p}$ denotes the expected equilibrium proposal:*

$$\bar{p} = \int_{t\in N} p_t d\mu(t).$$

[1] Suitable measurability conditions are that $p : N \to [0,1]$ be $\mathcal{A}$-measurable, and $\cap_{t\in N} A_t$ be Borel-measurable. When N is finite, no measurability conditions are needed.

[C] *The set A is a non–empty compact interval.*

PROOF: We must have $\mu(N^{\mathrm{a}}) > 0$, for otherwise $U_t(\sigma) = 0$ from Equation (4.1) for all $t \in N$, and citizens would have a profitable deviation by proposing any tax rate in $(0, 1)$. Let $\bar{p}^{\mathrm{a}}$ denote the average proposal of the citizens in N^{a}, that is

$$\bar{p}^{\mathrm{a}} = \frac{1}{\mu(N^{\mathrm{a}})} \int_{t \in N^{\mathrm{a}}} p_t d\mu(t).$$

This expression is well-defined since we have argued $\mu(N^{\mathrm{a}})$ to be strictly positive. We have for every $i \in N$,

$$v_i(\bar{p}^{\mathrm{a}}) \geq \frac{1}{\mu(N^{\mathrm{a}})} \int_{t \in N^{\mathrm{a}}} v_i(p_t) d\mu(t) = \frac{1 - \delta + \delta\mu(N^{\mathrm{a}})}{\mu(N^{\mathrm{a}})} U_i(\sigma) \geq U_i(\sigma),$$

where the first inequality follows from the concavity of v_i and the equality from Equation (4.1).

Now consider the interval $[\delta\bar{p}^{\mathrm{a}}, 1 - \delta + \delta\bar{p}^{\mathrm{a}}]$. Each of its points can be written in the form $(1 - \delta)\beta + \delta\bar{p}^{\mathrm{a}}$ for some $\beta \in [0, 1]$. Since for every $i \in N$,

$$v_i((1 - \delta)\beta + \delta\bar{p}^{\mathrm{a}}) \geq (1 - \delta)v_i(\beta) + \delta v_i(\bar{p}^{\mathrm{a}}) \geq \delta v_i(\bar{p}^{\mathrm{a}}) \geq \delta U_i(\sigma),$$

each point of the interval $[\delta\bar{p}^{\mathrm{a}}, 1 - \delta + \delta\bar{p}^{\mathrm{a}}]$ is unanimously accepted.

We conclude the proof of the theorem by demonstrating that each citizen in $N \setminus N^{\mathrm{a}}$ has a profitable deviation from σ, thereby obtaining a contradiction, and showing that $N^{\mathrm{a}} = N$. Thus take a citizen $t \in N$ such that p_t is not an element of A. Player t's equilibrium proposal p_t is rejected and leads to utility $\delta U_t(\sigma)$ for citizen t. To obtain the desired contradiction it is sufficient to show that there is a point $\beta \in A$ such that $v_t(\beta) > \delta U_t(\sigma)$, for then proposing β, rather than p_t, would be a profitable deviation for citizen t. Suppose first that $t \in N^- \cup N^+$. As we have seen above, $v_t(\beta) \geq \delta U_t(\sigma)$ for each β such that $\delta\bar{p}^{\mathrm{a}} \leq \beta \leq 1 - \delta + \delta\bar{p}^{\mathrm{a}}$. Since v_t is strictly increasing for $t \in N^-$ and strictly decreasing for $t \in N^+$ we must have $v_t(\beta) > \delta U_t(\sigma)$ whenever $\delta\bar{p}^{\mathrm{a}} < \beta < 1 - \delta + \delta\bar{p}^{\mathrm{a}}$. Suppose next $t \in N^0$. Then the function v_t is a positive constant, so it follows that any $\beta \in [0, 1]$, and in particular any $\beta \in A$, has the property that $v_t(\beta) > \delta v_t(\beta) \geq \delta U_t(\sigma)$.

From Equation (4.2) and the fact the function v_i is continuous and concave it follows that citizen i's acceptance set A_i is a compact interval. It follows that also the social acceptance set is a compact interval. □

Henceforth we let $[\beta^-, \beta^+]$ denote the interval A. By the preceding theorem the set A contains an interval of length $1 - \delta$, so it follows that

$$\beta^+ - \beta^- \geq 1 - \delta. \tag{4.3}$$

The proposal of a citizen of type t is the point in A closest to his most preferred point. Therefore, if $t \in N^+$ then $p_t = \beta^-$ and if $t \in N^-$ then $p_t = \beta^+$. For $t \in N^0$ the indirect utility function is a constant, so citizen t's proposal can be an arbitrary element of the social acceptance set. Hence Equation (4.1) simplifies to

$$U_i(\sigma) = m^- v_i(\beta^+) + m^+ v_i(\beta^-), \quad i \in N.$$

Thus the social acceptance set $[\beta^-, \beta^+]$ and the individual acceptance sets in an SSPEI are such that

$$[\beta^-, \beta^+] = \bigcap_{i \in N} A_i, \tag{4.4}$$

$$A_i = \{\beta \in [0,1] : v_i(\beta) \geq \delta[m^- v_i(\beta^+) + m^+ v_i(\beta^-)]\}, \quad i \in N. \tag{4.5}$$

In what follows we provide a characterization of the endpoints of the social acceptance set in an SSPEI as a fixed point of an appropriately defined function. Recall that h_i has been defined for each $i \in N^- \cup N^+$ as the inverse of the function v_i. For each $i \in N^- \cup N^+$ we define

$$f_{1i}(\beta^-) = h_i(\alpha^+ v_i(\beta^-)), \quad \text{where} \quad \alpha^+ = \delta m^+/(1 - \delta + \delta m^+), \tag{4.6}$$

$$f_{2i}(\beta^+) = h_i(\alpha^- v_i(\beta^+)), \quad \text{where} \quad \alpha^- = \delta m^-/(1 - \delta + \delta m^-). \tag{4.7}$$

Define the functions $f_1 : [0,1] \to [0,1]$ and $f_2 : [0,1] \to [0,1]$ by letting

$$f_1(\beta^-) = \min\big\{1, \inf_{i \in N^+} f_{1i}(\beta^-)\big\},$$
$$f_2(\beta^+) = \max\big\{0, \sup_{i \in N^-} f_{2i}(\beta^+)\big\}.$$

LEMMA 4.2: *Let $[\beta^-, \beta^+]$ be the social acceptance set induced by an SSPEI. Then*

$$\beta^+ = f_1(\beta^-),$$
$$\beta^- = f_2(\beta^+).$$

PROOF: We show that the second of these equations holds. The proof of the first one is similar. Since β^- is in the social acceptance set, it belongs to every citizen's individual acceptance set A_i. Hence by (4.5)

$$v_i(\beta^-) \geq \delta m^- v_i(\beta^+) + \delta m^+ v_i(\beta^-), \quad i \in N.$$

Using the fact that $m^- + m^+ = 1$ and rearranging we find that

$$v_i(\beta^-) \geq \alpha^- v_i(\beta^+), \quad i \in N.$$

Now consider the citizens $i \in N^-$. For each such citizen i the function h_i is strictly increasing. Hence applying h_i to the preceding equation yields

$$\beta^- = h_i(v_i(\beta^-)) \geq h_i(\alpha^- v_i(\beta^+)) = f_{2i}(\beta^+), \quad i \in N^-.$$

We conclude that $\beta^- \geq f_2(\beta^+)$. Now suppose $\beta^- \geq f_2(\beta^+) + \varepsilon$ for some $\varepsilon > 0$. Then

$$\beta^- - \varepsilon \geq h_i(\alpha^- v_i(\beta^+)), \quad i \in N^-.$$

Applying the increasing function v_i to the above inequality yields

$$v_i(\beta^- - \varepsilon) \geq v_i(h_i(\alpha^- v_i(\beta^+))) = \alpha^- v_i(\beta^+), \ i \in N^-.$$

But then the point

$$\beta = (1 - \delta m^+)(\beta^- - \varepsilon) + \delta m^+ \beta^-$$

is in the acceptance set of each citizen $i \in N^-$, because

$$\begin{aligned} v_i(\beta) &\geq (1 - \delta m^+) v_i(\beta^- - \varepsilon) + \delta m^+ v_i(\beta^-) \\ &\geq (1 - \delta m^+) \alpha^- v_i(\beta^+) + \delta m^+ v_i(\beta^-) \\ &= \delta m^- v_i(\beta^+) + \delta m^+ v_i(\beta^-) \\ &= \delta U_i(\sigma), \end{aligned}$$

where the inequality in the first line follows by concavity of v_i. Since $\beta < \beta^-$ is clearly in the acceptance set of each citizen in $i \in N^0 \cup N^+$, it is in the social acceptance set. This clearly contradicts the fact that β^- is the left endpoint of the social acceptance set. □

The first equality in the preceding lemma expresses that β^+ should be sufficiently low to make it acceptable for all above average income citizens, and similarly the second equality specifies that β^- should be sufficiently high to be acceptable for all below average income citizens. It motivates the following definition.

Definition 4.3: A *tax equilibrium* is pair (β^-, β^+) satisfying

$$\beta^+ = f_1(\beta^-),$$
$$\beta^- = f_2(\beta^+).$$

As is immediate from the definition, every SSPEI induces a tax equilibrium. We now establish the converse, stating every tax equilibrium corresponds to an SSPEI, showing that the concepts are equivalent.

THEOREM 4.4: *Given a tax equilibrium (β^-, β^+) there is an SSPEI σ with social acceptance set equal to $[\beta^-, \beta^+]$.*

PROOF: Define p_t to be β^+ for each $t \in N^-$ and β^- for each $t \in N^+$. Let p_t be an arbitrary point of $[0,1]$ for each $t \in N^0$. Define the individual acceptance set by Equation (4.5). An argument similar to that in the proof of Lemma 4.2 can be used to show that Equation (4.4) holds.

It is a routine excercise to demonstrate that no type has a profitable one-shot deviation from σ. Here, a one-shot deviation in a subgame is a single deviation by a player at the root of the subgame. It follows from a standard argument, see for instance Fudenberg and Tirole (1991), that if there is a subgame where a type has some profitable deviation from σ, then there must also be a subgame where this type has a profitable one-shot deviation. □

Define the function $f : [0,1]^2 \to [0,1]^2$ by letting

$$f(\beta^+, \beta^-) = (f_1(\beta^-), f_2(\beta^+)).$$

Clearly the tax equilibria are exactly the fixed points of the function f.

THEOREM 4.5: *A tax equilibrium exists.*

PROOF: The set $[0,1]^2$ is a complete lattice and the function f has the property that $f_1(\beta) \leq f_1(\beta')$ and $f_2(\beta) \leq f_2(\beta')$ whenever $\beta \leq \beta'$. Hence Tarski's fixed point theorem implies that f has a fixed point. □

THEOREM 4.6: *A tax equilibrium is unique.*

PROOF: Consider the function $g : [0,1] \to \mathbb{R}$ defined by $g(\beta) = f_1(f_2(\beta)) - \beta$. Obviously, a zero point β of g is in a one to one relationship with a tax equilibrium $(f_2(\beta), \beta)$ and it follows from the previous paragraph that g has at least one zero point. We argue that g is strictly decreasing, thereby showing that it has a unique zero point.

We write

$$g(\beta) = f_1(f_2(\beta)) - f_2(\beta) + f_2(\beta) - \beta.$$

We show first that $f_2(\beta) - \beta$ is strictly decreasing in β. Consider some $i \in N^-$ and β, β' with $\beta < \beta'$. Since h_i is a convex and increasing function, we have

$$\begin{aligned} f_{2i}(\beta) - f_{2i}(\beta') &= h_i(\alpha^- v_i(\beta)) - h_i(\alpha^- v_i(\beta')) \\ &\geq \alpha^-[h_i(v_i(\beta)) - h_i(v_i(\beta'))] \\ &= \alpha^-(\beta - \beta'), \end{aligned}$$

hence

$$f_{2i}(\beta) \geq \alpha^-(\beta - \beta') + f_{2i}(\beta').$$

Taking the supremum with respect to $i \in N^-$ on both sides of the preceding equation yields

$$\sup_{i \in N^-} f_{2i}(\beta) \geq \alpha^-[\beta - \beta'] + \sup_{i \in N^-} f_{2i}(\beta').$$

Hence

$$\sup_{i \in N^-} f_{2i}(\beta) - \beta \geq \sup_{i \in N^-} f_{2i}(\beta') - \beta' + (1 - \alpha^-)(\beta' - \beta) > \sup_{i \in N^-} f_{2i}(\beta') - \beta',$$

which implies that

$$\sup_{i \in N^-} f_{2i}(\beta) - \beta$$

is strictly decreasing in β. Notice that

$$f_2(\beta) - \beta = \max\big\{ -\beta, \sup_{i \in N^-} f_{2i}(\beta) - \beta \big\},$$

so it is the maximum of two expressions strictly decreasing in β, and therefore strictly decreasing itself. A completely symmetric argument shows that $f_1(\beta) - \beta$ is strictly decreasing in β. Since the function f_2 is increasing, $f_1(f_2(\beta)) - f_2(\beta)$ is decreasing in β. Since $f_2(\beta) - \beta$ is strictly decreasing in β, we have that g is a strictly decreasing function. □

The unanimous approval procedure results in a unique tax equilibrium (β^-, β^+) with $\beta^- < \beta^+$. The tax rate proposed by an above average income citizen is β^-, whereas β^+ is proposed by below average income citizens. A proposal β strictly smaller than β^- would be rejected by at least one below average income citizen, and a proposal β strictly greater than β^+ by at least one above average income citizen. At equilibrium rejections do not occur.

5 Boldness

We show that, along any convergent sequence of tax equilibria, as the discount factor converges to one, the social acceptance set converges to a singleton set.

LEMMA 5.1: *Let $(\delta_n)_{n \in \mathbb{N}}$ be a sequence converging to 1 and, for $n \in \mathbb{N}$, let (β_n^-, β_n^+) be the tax equilibrium corresponding to discount factor δ_n. Suppose the sequence $(\beta_n^-)_{n \in \mathbb{N}}$ converges to a point β^- and the sequence $(\beta_n^+)_{n \in \mathbb{N}}$ converges to β^+. Then $\beta^- = \beta^+$.*

PROOF: We know that $\beta_n^- \leq \beta_n^+$ for each n. Hence $\beta^- \leq \beta^+$. The point β_n^-, being an element of the social acceptance set in a tax equilibrium, is accepted by all citizens. In particular, for $i \in N^-$ we have $v_i(\beta_n^-) \geq \delta_n(m^- v_i(\beta_n^+) + m^+ v_i(\beta_n^-))$. Taking the limit of both sides of the inequality as n goes to infinity we obtain $v_i(\beta^-) \geq m^- v_i(\beta^+) + m^+ v_i(\beta^-)$. Rearranging yields $v_i(\beta^-) \geq v_i(\beta^+)$. Since v_i is an increasing function for $i \in N^-$ we have $\beta^- \geq \beta^+$. □

We define a procedurally fair tax rate as the limit of proposal in a tax equilibrium when δ converges to one.

DEFINITION 5.2: The tax rate β is *procedurally fair* if it is the limit of a sequence $(\beta_{t,n})_{n\in\mathbb{N}}$, where $\beta_{t,n}$ is the proposal of a citizen t in the tax equilibrium corresponding to δ_n and $(\delta_n)_{n\in\mathbb{N}}$ is a sequence converging to one.

In this section we provide a simple characterization of the procedurally fair tax rate. In the characterization, the concept of boldness as introduced in Aumann and Kurz (1977) plays an important role. The boldness of a citizen of type t at consumption c_t is defined to be the quantity $u_t'(c_t)/u_t(c_t)$.

Consider a gamble where a citizen of type t receives ℓ_t with probability q_t and $c_t + \varepsilon$ with probability $1 - q_t$, where $\varepsilon > 0$. Let $q_t(c_t, \varepsilon)$ be the maximum probability for which a citizen of type t weakly prefers the gamble over consuming c_t for sure. As pointed out in Roth (1989), boldness corresponds to the maximum probability for which type t is willing to accept the gamble, per dollar of additional gains, when ε tends to zero. That is, $b_t(c_t) = \lim_{\varepsilon\downarrow 0} q_t(c_t, \varepsilon)/\varepsilon$. Aumann and Kurz (1977) observe that the Nash bargaining solution for the two player case can be characterized as selecting a division of the potential surplus at which the players are equally bold.

Since we do not assume that utility functions are differentiable, we extend the notion of boldness in the following way. The *boldness* of a citizen of type t at $c_t > \ell_t$ is defined as

$$b_t(c_t) = \partial_+ u_t(c_t)/u_t(c_t),$$

where $\partial_+ u_t$ denotes the right derivative of u_t. Left derivatives will be denoted by ∂_-. We define $b_t(\ell_t) = +\infty$. Since utility functions are assumed to be concave, the right derivative of the utility function is well-defined. Moreover, boldness is a strictly decreasing function of consumption, as the numerator in the definition of boldness is decreasing by concavity of u_t and the denominator is strictly increasing since u_t is strictly increasing in consumption.

The next result shows that boldness still admits an interpretation in terms of gambles.

THEOREM 5.3: *For $c_t > \ell_t$ it holds that $b_t(c_t) = \lim_{\varepsilon\downarrow 0} q_t(c_t, \varepsilon)/\varepsilon$.*

PROOF: By continuity of u_t and for ε sufficiently small it holds that

$$u_t(c_t) = q_t(c_t, \varepsilon)u_t(\ell_t) + (1 - q_t(c_t, \varepsilon))u_t(c_t + \varepsilon).$$

We use that $u_t(\ell_t) = 0$ and rearrange terms to find that

$$\frac{q_t(c_t, \varepsilon)}{\varepsilon} = \frac{u_t(c_t + \varepsilon) - u_t(c_t)}{\varepsilon} \frac{1}{u_t(c_t + \varepsilon)}.$$

When ε tends to zero, we find that the first term in the product converges to the right derivative of u_t at c_t due to concavity of u_t and the second term converges to $u_t(c_t)$ because of continuity, so

$$\lim_{\varepsilon \downarrow 0} \frac{q_t(c_t, \varepsilon)}{\varepsilon} = \frac{\partial_+ u_t(c_t)}{u_t(c_t)}.$$

□

It follows immediately from Theorem 5.3 and the concavity of u_t that an alternative characterization of $b_t(c_t)$ can be given as the limit inferior of difference quotients of u_t at c_t. We observe that b_t is a strictly decreasing function, which may have points of discontinuity.

For each $\beta \in (0, 1)$ define

$$\begin{aligned} d_{1t}(\beta) &= -\frac{\partial_- v_t(\beta)}{v_t(\beta)} = (w_t - \bar{w}) b_t(w_t + \beta(\bar{w} - w_t)), \\ d_{2t}(\beta) &= \frac{\partial_+ v_t(\beta)}{v_t(\beta)} = (\bar{w} - w_t) b_t(w_t + \beta(\bar{w} - w_t)). \end{aligned}$$

The above expressions are well–defined since $v_t(\beta) > 0$ for each $\beta \in (0, 1)$.

We will show that these functions yield the appropriate measure of boldness of a citizen of type t at a tax rate equal to β, where d_{1t} applies to types with above average income and d_{2t} to those with below average income. We refer to this measure of boldness as *tax rate boldness*. Indeed, for a type t in N^+, tax rate boldness $d_{1t}(\beta)$ is equal to the usual definition of boldness applied to the (increasing) indirect utility function $-v_t$, which in turn is equal to his boldness at the consumption $w_t + \beta(\bar{w} - w_t)$ induced by the tax rate β, multiplied by $w_t - \bar{w}$, which equals the marginal change in consumption due to a change in the tax rate. Similarly, $d_{2t}(\beta)$ is the appropriate measure for boldness corresponding to tax rate proposals for below average income types.

Since boldness is strictly decreasing in consumption, it holds that tax rate boldness d_{1t} is strictly increasing in β, whereas d_{2t} is strictly decreasing in β. What will matter in the end is to convince the boldest citizen to accept a proposal, which motivates the following definitions. For $\beta \in (0, 1)$ let

$$\begin{aligned} d_1(\beta) &= \sup_{t \in N^+} d_{1t}(\beta), \\ d_2(\beta) &= \sup_{t \in N^-} d_{2t}(\beta). \end{aligned}$$

Lemmas 5.4 and 5.5 present some crucial properties of the functions d_1 and d_2.

LEMMA 5.4: *The function d_1 is a strictly increasing function on $(0,1)$ assuming only finite values.*

PROOF: Consider some $t \in N^+$ and some $\beta \in (0,1)$. Using $v_t(1) \geq 0$ and the concavity of v_t, we find that

$$-(1-\beta)\partial_- v_t(\beta) \leq v_t(\beta).$$

Rearranging yields the inequality

$$d_{1t}(\beta) \leq \frac{1}{1-\beta}.$$

This establishes the finiteness of d_1 on $(0,1)$. We show that d_1 is strictly increasing. Let β and β' be elements of $(0,1)$ such that $\beta < \beta'$. For $t \in N^+$ we have

$$-\partial_- v_t(\beta)(\beta' - \beta) \leq v_t(\beta) - v_t(\beta')$$

We therefore obtain

$$\begin{aligned} -\partial_- v_t(\beta)(v_t(\beta') - \partial_- v_t(\beta')(\beta' - \beta)) &\leq -\partial_- v_t(\beta) v_t(\beta') - \partial_- v_t(\beta')(v_t(\beta) - v_t(\beta')) \\ &= (\partial_- v_t(\beta') - \partial_- v_t(\beta)) v_t(\beta') - \partial_- v_t(\beta') v_t(\beta) \\ &\leq -\partial_- v_t(\beta') v_t(\beta), \end{aligned}$$

where the last inequality uses $v_t(\beta') \geq 0$ and concavity. Dividing both sides of the inequality by $v_t(\beta) v_t(\beta')$ gives

$$d_{1t}(\beta)(1 + (\beta' - \beta) d_{1t}(\beta')) \leq d_{1t}(\beta').$$

Dividing by $d_{1t}(\beta')$ yields the inequality

$$d_{1t}(\beta) \leq \frac{d_{1t}(\beta')}{1 + (\beta' - \beta) d_{1t}(\beta')}.$$

Taking the supremum with respect to $t \in N^+$ yields

$$d_1(\beta) = \sup_{t \in N^+} d_{1t}(\beta) \leq \sup_{t \in N^+} \frac{d_{1t}(\beta')}{1 + (\beta' - \beta) d_{1t}(\beta')} \leq \frac{d_1(\beta')}{1 + (\beta' - \beta) d_1(\beta')} < d_1(\beta').$$

This completes the proof. □

We now show that the function d_2 is strictly decreasing exploiting a symmetry argument. This construction will be used once more in the proof of Theorem 6.1. Starting from given a society we consider a new society where the citizens' incomes are "flipped over" the average income. The old and the new societies are "strategically equivalent" in the sense that a tax rate of β in the new society corresponds to a tax rate of $1 - \beta$ in the original society. In order to make use of this argument, we have allowed for negative pre–tax incomes.

LEMMA 5.5: *The function d_2 is a strictly decreasing function on $(0,1)$ assuming only finite values.*

PROOF: From a given society $S = (N, \mu, w_t, u_t)_{t\in N}$ we create the society $\dot{S} = (N, \mu, \dot{w}_t, \dot{u}_t)_{t\in N}$ where $\dot{w}_t = 2\bar{w} - w_t$ and $\dot{u}_t(c_t) = u_t(w_t - \bar{w} + c_t)$, for $c_t \geq \dot{\ell}_t = \ell_t + w_t - \bar{w}$. The average income in the society $\dot{S}$ equals that in the society S, namely $\bar{w}$. The citizens $\dot{N}^-$ with below–average income in $\dot{S}$ are precisely those with above–average income in the society S and conversely, the citizens $\dot{N}^+$ with above–average income in $\dot{S}$ are those with below–average income in S. Thus $\dot{N}^- = N^+$ and $\dot{N}^+ = N^-$. The indirect utility functions $\dot{v}_t$ in the society $\dot{S}$ are related to those in the society S by

$$\dot{v}_t(\beta) = v_t(1-\beta), \quad \beta \in [0,1]. \tag{5.8}$$

It follows that, if (β^-, β^+) is a tax equilibrium in the society S, then $(1-\beta^+, 1-\beta^-)$ is a tax equilibrium in the society $\dot{S}$.

For each $\beta \in (0,1)$ we have

$$\partial_- \dot{v}_t(\beta) = -\partial_+ v_t(1-\beta) \text{ and } \partial_+ \dot{v}_t(\beta) = -\partial_- v_t(1-\beta).$$

It follows that

$$\dot{d}_1(\beta) = d_2(1-\beta) \text{ and } \dot{d}_2(\beta) = d_1(1-\beta). \tag{5.9}$$

The previous lemma, when applied to the society $\dot{S}$, shows that $\dot{d}_1$ is a strictly increasing function that assumes only finite values. There result follows since $d_2(\beta) = \dot{d}_1(1-\beta)$. □

For each $\beta \in (0,1)$ we define

$$z(\beta) = m^- d_2(\beta) - m^+ d_1(\beta).$$

It will be convenient to extend the function z to all of the interval $[0,1]$ by letting $z(0) = +\infty$ and $z(1) = -\infty$. The function z is the difference between the supremum of tax rate boldness among below average income types multiplied by their probability mass m^- and the supremum of tax rate boldness among above average income types multiplied by their probability mass m^+. The function z can be interpreted as an "excess bargaining power function," where positive values of $z(\beta)$ indicate that below average income types have more bargaining power than above average income types, causing increases in the tax rate, and the reverse when $z(\beta)$ is negative. Indeed, $m^+ d_1(\beta)$ represents the bargaining power of the rich and $m^- d_2(\beta)$ the bargaining power of the poor. We can think of $z(\beta)$ as the direction and magnitude by which β would be adjusted as a consequence of the prevailing bargaining forces.

Lemmas 5.4 and 5.5 yield the following result.

LEMMA 5.6: *The function z is strictly decreasing on* $[0,1]$.

A point $\beta^* \in [0,1]$ is called a *generalized zero point* of z if $z(\beta) > 0$ for $\beta < \beta^*$ and $z(\beta) < 0$ for $\beta > \beta^*$. When applied to the case $\beta^* = 0$, then being a generalized zero point means $z(\beta) < 0$ for all $\beta > 0$, and similarly $\beta^* = 1$ is a generalized zero point of z if $z(\beta) > 0$ for all $\beta < 1$. Since z is strictly decreasing by Lemma 5.6, it follows that z has a unique generalized zero point.

COROLLARY 5.7: *The function z has a unique generalized zero point.*

At the generalized zero point β^* of z the difference between $m^- d_2(\beta)$ and $m^+ d_1(\beta)$ is minimized, and is equal to zero if z is continuous at β^*. The point β^* is therefore the tax rate where the bargaining power of citizens with below average income is as close as possible to that of above average income citizens. We will show in Theorem 6.5 that all proposals made in bargaining equilibria converge to the unique generalized zero point of z when δ converges to 1.

6 Procedurally Fair Taxation

In this section we establish that the generalized zero point of the function z lies in the social acceptance set of the tax equilibrium, for each value of the discount factor.

THEOREM 6.1: *The tax equilibrium (β^-, β^+) satisfies $\beta^- \leq \beta^* \leq \beta^+$, where β^* is the generalized zero point of z.*

At this point it is useful to provide some intuition for why the tax rate boldness terms d_1 and d_2 appear in our characterizations. Suppose indirect utility functions are differentiable and suppose that we can replace the indirect utility functions by their first-order approximations around the expected equilibrium offer $\bar{\beta}$:

$$v_t(\beta) \approx v_t(\bar{\beta}) + v_t'(\bar{\beta})(\beta - \bar{\beta}).$$

A type $t \in N$ accepts a proposal β if and only if $v_t(\bar{\beta}) + v_t'(\bar{\beta})(\beta - \bar{\beta}) \geq \delta v_t(\bar{\beta})$. Using that for $t \in N^+$, $v_t'(\bar{\beta}) < 0$ and for $t \in N^-$, $v_t'(\bar{\beta}) > 0$, we find that a type $t \in N^+$ accepts all proposals β satisfying

$$\beta \leq \bar{\beta} - (1-\delta)\frac{v_t(\bar{\beta})}{v_t'(\bar{\beta})},$$

whereas types $t \in N^-$ accept all proposals β satisfying

$$\beta \geq \bar{\beta} - (1-\delta)\frac{v_t(\bar{\beta})}{v_t'(\bar{\beta})}.$$

The endpoints of the social acceptance set are therefore given by

$$\beta^- = \bar{\beta} - (1-\delta) \inf_{t \in N^-} \left\{ \frac{v_t(\bar{\beta})}{v_t'(\bar{\beta})} \right\} = \bar{\beta} - \frac{1-\delta}{d_2(\bar{\beta})},$$

$$\beta^+ = \bar{\beta} - (1-\delta) \sup_{t \in N^+} \left\{ \frac{v_t(\bar{\beta})}{v_t'(\bar{\beta})} \right\} = \bar{\beta} + \frac{1-\delta}{d_1(\bar{\beta})}.$$

Since the expected equilibrium offer is given by $\bar{\beta} = m^+\beta^- + m^-\beta^+$, we find that

$$\bar{\beta} = m^+\bar{\beta} - m^+\frac{1-\delta}{d_2(\bar{\beta})} + m^-\bar{\beta} + m^-\frac{1-\delta}{d_1(\bar{\beta})}.$$

Since $m^+ + m^- = 1$, we can rearrange terms and find that $m^+ d_1(\bar{\beta}) = m^- d_2(\bar{\beta})$. In other words, we have found that the expected proposal is a zero point of the function z.

Since first-order approximations are not exact, it is not exactly true that the expected proposal in the tax equilibrium equals the zero point β^* of z, even if all indirect utility functions were differentiable. Theorem 6.1 demonstrates that nevertheless it is true in general that $\beta^- \leq \beta^* \leq \beta^+$. It follows from Theorem 6.1 that the procedurally fair tax rate is unique and is equal to the generalized zero point of the function z.

The proof of Theorem 6.1 consists of Lemmas 6.2–6.4 below.

Consider the bargaining equilibrium (β^-, β^+) and consider the bargaining power of the poor and the rich, respectively, at the upper bound β^+ of the social acceptance set. Lemma 6.2 states that the size of the social acceptance set $\beta^+ - \beta^-$ multiplied by the bargaining power of the poor $m^- d_2(\beta^+)$ is bounded from above by $(1-\delta)/\delta$, and is greater than or equal to this number when multiplied by the bargaining power $m^+ d_1(\beta^-)$ of the rich.

LEMMA 6.2: *Consider the tax equilibrium (β^-, β^+). If $\beta^+ < 1$, then*

1. $\delta m^- d_2(\beta^+)(\beta^+ - \beta^-) \leq (1-\delta)$.
2. $\delta m^+ d_1(\beta^+)(\beta^+ - \beta^-) \geq (1-\delta)$.

PROOF: For $t \in N^-$, it holds that the proposal β^- is accepted, so $\delta(m^- v_t(\beta^+) + m^+ v_t(\beta^-)) \leq v_t(\beta^-)$. Rewriting this inequality results in

$$\delta m^-(v_t(\beta^+) - v_t(\beta^-)) \leq (1-\delta)v_t(\beta^-) \leq (1-\delta)v_t(\beta^+).$$

Since v_t is concave, we find that

$$\delta m^- \partial_+ v_t(\beta^+)(\beta^+ - \beta^-) \leq (1-\delta)v_t(\beta^+),$$

and therefore $\delta m^- d_{2t}(\beta^+)(\beta^+ - \beta^-) \leq 1-\delta$. Since this inequality holds for all $t \in N^-$, we obtain

$$\delta m^- d_2(\beta^+)(\beta^+ - \beta^-) = \sup_{t \in N^-} \delta m^- d_{2t}(\beta^+)(\beta^+ - \beta^-) \leq 1-\delta.$$

This proves Lemma 6.2.1.

Consider some $t \in N^+$. By concavity of v_t we have that $-\partial_- v_t(\beta^+)(\beta^+ - \beta^-) \geq v_t(\beta^-) - v_t(\beta^+)$. It then follows that

$$d_{1t}(\beta^+)(\beta^+ - \beta^-) \geq \frac{v_t(\beta^-)}{v_t(\beta^+)} - 1.$$

We take the supremum over all $t \in N^+$ and find that

$$d_1(\beta^+)(\beta^+ - \beta^-) \geq \sup_{t \in N^+} \frac{v_t(\beta^-)}{v_t(\beta^+)} - 1.$$

We complete the proof of Lemma 6.2.2 by showing that

$$\delta m^+ \left(\sup_{t \in N^+} \frac{v_t(\beta^-)}{v_t(\beta^+)} - 1 \right) = 1 - \delta,$$

or equivalently

$$\sup_{t \in N^+} \frac{\delta(m^- v_t(\beta^+) + m^+ v_t(\beta^-))}{v_t(\beta^+)} = 1.$$

Since all $t \in N^+$ accept the proposal β^+, we have

$$\frac{\delta(m^- v_t(\beta^+) + m^+ v_t(\beta^-))}{v_t(\beta^+)} \leq 1, \quad t \in N^+.$$

Suppose there is an $\varepsilon > 0$ with the property

$$\frac{\delta(m^- v_t(\beta^+) + m^+ v_t(\beta^-))}{v_t(\beta^+)} \leq 1 - \varepsilon$$

for all $t \in N^+$. Then

$$\begin{aligned} v_t((1-\varepsilon)\beta^+ + \varepsilon) &\geq (1-\varepsilon)v_t(\beta^+) + \varepsilon v_t(1) \geq (1-\varepsilon)v_t(\beta^+) \\ &\geq \delta(m^- v_t(\beta^+) + m^+ v_t(\beta^-)). \end{aligned}$$

All citizens $t \in N^+$ (as well as citizens in $N^- \cup N^0$) therefore accept the proposal $(1-\varepsilon)\beta^+ + \varepsilon > \beta^+$ thereby contradicting that β^+ is the upper bound of the social acceptance set. □

We are now in a position to prove the first half of Theorem 6.1.

LEMMA 6.3: *The tax equilibrium (β^-, β^+) satisfies $\beta^* \leq \beta^+$, where β^* is the generalized zero point of z.*

PROOF: The result is obviously true when $\beta^+ = 1$. In particular, if $\delta = 0$ then $\beta^+ = 1$ and $\beta^- = 0$ by Equation (4.3). Suppose now $\delta > 0$ and $\beta^+ < 1$. Subtracting the second inequality of Lemma 6.2 from the first one gives $\delta z(\beta^+)(\beta^+ - \beta^-) \leq 0$. Since by (4.3) $0 < \beta^+ - \beta^-$, it follows that $z(\beta^+) \leq 0$. The result follows since z is a decreasing function. □

We prove the second half of Theorem 6.1 using the already familiar symmetry argument.

LEMMA 6.4: *The tax equilibrium (β^-, β^+) satisfies $\beta^- \leq \beta^*$, where β^* is the generalized zero point of z.*

PROOF: Consider the society $\dot{S}$ as defined in the proof of Lemma 5.5. By Equation (5.9) the characteristic function $\dot{z}$ in $\dot{S}$ is related to that in S by $\dot{z}(\beta) = -z(1-\beta)$, $\beta \in [0,1]$. In particular, the point $1 - \beta^*$ is the generalized zero point of $\dot{z}$. It is clear from Equation (5.8) that, if (β^-, β^+) is a tax equilibrium in the society S, then $(1-\beta^+, 1-\beta^-)$ is a tax equilibrium in the society $\dot{S}$. Lemma 6.3, when applied to the society $\dot{S}$, shows that $1 - \beta^* \leq 1 - \beta^-$ and the result follows. □

Combining Lemma 5.1 with Theorem 6.1 we obtain the following result.

THEOREM 6.5: *Let $(\delta_n)_{n\in\mathbb{N}}$ be a sequence converging to 1 and, for $n \in \mathbb{N}$, let (β_n^-, β_n^+) be the tax equilibrium of the game corresponding to δ_n. Then both sequences $(\beta_n^-)_{n\in\mathbb{N}}$ and $(\beta_n^+)_{n\in\mathbb{N}}$ converge to the generalized zero point β^* of z.*

COROLLARY 6.6: *A society has a unique procedurally fair tax rate.*

As an illustration, consider the case where all citizens $t \in N$ have constant relative risk aversion preferences

$$u_t(c_t) = c_t^{1-\gamma}, \quad c_t \in \mathbb{R}_+,$$

where $\gamma \in [0,1)$ is the coefficient of relative risk aversion. Moreover, we assume that

$$\inf_{t\in N} w_t = 0 \text{ and } \sup_{t\in N} w_t = +\infty.$$

It is straightforward to derive that tax rate boldness of an above average income citizen $t \in N^+$ is given by

$$d_{1t}(\beta) = \frac{(1-\gamma)(w_t - \bar{w})}{(1-\beta)w_t + \beta\bar{w}}, \quad \beta \in [0,1].$$

Since, for given tax rate β, this expression is increasing in w_t, it attains the maximum value when t is equal to the highest income citizen. Since the support of μ is unbounded we have that the bargaining power of the rich is given by

$$m^+ d_1(\beta) = m^+ \sup_{t\in N^+} d_{1t}(\beta) = m^+ \frac{1-\gamma}{1-\beta}, \quad \beta \in [0,1).$$

The tax rate boldness of a below average income citizen $t \in N^-$ is equal to

$$d_{2t}(\beta) = \frac{(1-\gamma)(\bar{w} - w_t)}{(1-\beta)w_t + \beta\bar{w}}, \quad \beta \in (0,1].$$

For given tax rate β, this expression is decreasing in w_t, and the maximum over $t \in N^-$ is therefore attained by the lowest income citizen. We have that the bargaining power of the poor is equal to

$$m^- d_2(\beta) = m^- \sup_{t \in N^-} d_{2t}(\beta) = m^- \frac{1-\gamma}{\beta}, \quad \beta \in (0,1].$$

The unique zero point of the excess bargaining power function

$$z(\beta) = m^- d_2(\beta) - m^+ d_1(\beta) = m^- \frac{1-\gamma}{\beta} - m^+ \frac{1-\gamma}{1-\beta}, \quad \beta \in (0,1),$$

is given by $\beta = m^-$. The unique procedurally fair tax rate of a society populated by constant relative risk aversion citizens is equal to the probability mass of below average income citizens, irrespective of the shape of the distribution μ of types within the population.

7 Heterogeneous Societies

In this section we further analyze the procedurally fair tax rate for societies with sufficiently heterogeneous citizens. What we have in mind is that there is sufficient dispersion in preferences among citizens, in particular in terms of boldness. From a technical point of view, we make the assumption that there are two risk–neutral types, one having the lowest pre–tax income level and the other one the highest. Let

$$w^- = \inf_{t \in N} w_t \text{ and } w^+ = \sup_{t \in N} w_t.$$

Throughout this section we make the following assumptions.

Assumption 7.1: It holds that $w^- \geq 0$ and $w^+ < +\infty$.

Assumption 7.2: For each $t \in N$ the utility function $u_t : [0, +\infty) \to \mathbb{R}$ is continuous, concave, strictly increasing, and $u_t(0) = 0$.

Assumption 7.3: There exists a type $a \in N$ with the property that $w_a = w^-$ and $u_a(c_a) = c_a$, $c_a \geq 0$. There exists a type $b \in N$ with the property that $w_b = w^+$ and $u_b(c_b) = c_b$, $c_b \geq 0$.

We define

$$\beta^* = m^- \frac{w^+}{w^+ - \bar{w}} - m^+ \frac{w^-}{\bar{w} - w^-},$$

$$\beta^{**} = \begin{cases} 0, & \text{if } \beta^* \leq 0, \\ \beta^*, & \text{if } 0 \leq \beta^* \leq 1, \\ 1, & \text{if } 1 \leq \beta^*. \end{cases}$$

The main result of this section states that the tax rate β^{**} is the procedurally fair tax rate of any society satisfying the above assumptions. We begin our analysis by considering a society consisting of only two citizens satisfying the conditions of Assumption 7.3. More precisely, given a society S satisfying the maintained assumptions consider the society T consisting of only the types a and b. Let the probability mass for the types a and b be m^- and m^+, respectively. We first show β^{**} to be the procedurally fair tax rate for the society T. Then we argue that every tax equilibrium of the society S, for each value of the discount factor, equals that of the society T.

LEMMA 7.4: *Let (β^-, β^+) be the tax equilibrium of the society T. Then*

[A] $\beta^- \leq \beta^{**} \leq \beta^+$.

[B] *If $\beta^- > 0$ and $\beta^+ < 1$ then $\beta^* = \beta^{**}$ and is equal to the expected equilibrium proposal.*

PROOF: For types $t = a, b$ the indirect utility function v_t is given by $v_t(\beta) = w_t + \beta(\bar{w} - w_t)$, $\beta \in [0, 1]$, the function h_t by $h_t(v) = (w_t - v)/(w_t - \bar{w})$, $v \in [0, +\infty)$. Using the definitions (4.6)–(4.7) it is straightforward to compute the functions f_{1b} and f_{2a}:

$$\begin{aligned} f_{1b}(\beta^-) &= \frac{1-\delta}{1-\delta+\delta m^+}\frac{w^+}{w^+ - \bar{w}} + \beta^- \frac{\delta m^+}{1-\delta+\delta m^+}. \\ f_{2a}(\beta^-) &= \frac{1-\delta}{1-\delta+\delta m^-}\frac{w^-}{w^- - \bar{w}} + \beta^+ \frac{\delta m^-}{1-\delta+\delta m^-}. \end{aligned}$$

Now suppose (β^-, β^+) is a tax equilibrium. Then, by definition it holds that

$$\beta^+ \leq f_{1b}(\beta^-) \quad \text{with equality if} \quad \beta^+ < 1, \tag{7.1}$$

$$\beta^- \geq f_{2a}(\beta^+) \quad \text{with equality if} \quad \beta^- > 0. \tag{7.2}$$

To prove claim [B] suppose $\beta^- > 0$ and $\beta^+ < 1$. In this case both (7.1) and (7.2) hold with equality. Multiplying the first of these by $m^-(1 - \delta + \delta m^+)$, the second one by $m^+(1 - \delta + \delta m^-)$, and adding them up we obtain

$$\begin{aligned} (1-\delta)\bar{\beta} + \delta m^- m^+(\beta^- + \beta^+) &= m^-(1-\delta+\delta m^+)\beta^+ + m^+(1-\delta+\delta m^-)\beta^- \\ &= m^-(1-\delta+\delta m^+) f_{1b}(\beta^-) + m^+(1-\delta+\delta m^-) f_{2a}(\beta^+) \\ &= (1-\delta)\beta^* + \delta m^- m^+(\beta^- + \beta^+). \end{aligned}$$

It follows that $\bar{\beta} = \beta^*$. Hence $0 \leq \beta^* \leq 1$ so that $\beta^{**} = \beta^*$. This proves claim [B].

To prove claim [A] suppose that $\beta^{**} < \beta^-$. Then in particular $\beta^{**} < 1$ implying that $\beta^* \leq \beta^{**}$. Since $\beta^- < \bar{\beta}$ we obtain the inequality $\beta^* < \bar{\beta}$. On the other hand, since $0 < \beta^-$ the inequality in (7.2) holds with equality. Multiplying (7.1) by $m^-(1-\delta+\delta m^+)$ and (7.2) by $m^+(1-\delta+\delta m^-)$, adding them up and rearranging the result yields $\bar{\beta} \leq \beta^*$, contrary to the earlier conclusion. Similarly one shows that $\beta^+ < \beta^{**}$ is impossible. □

As an immediate corollary we obtain the following lemma.

LEMMA 7.5: *The tax rate β^{**} is the procedurally fair tax rate of the society T.*

THEOREM 7.6: *The pair (β^-, β^+) is a tax equilibrium of the society S if and only if it is a tax equilibrium of the society T.*

We prove the theorem by showing that the function f as defined in the preceding section for the society S equals that for the society T.

LEMMA 7.7: *The function f for the society S equals that for the society T.*

PROOF: We prove that for each $\beta \in [0,1]$ it holds that

$$\inf_{i \in N^+} f_{1i}(\beta) = f_{1b}(\beta),$$
$$\sup_{i \in N^-} f_{2i}(\beta) = f_{2a}(\beta).$$

We prove the first equality. The proof of the second equality is similar.

Fix some $\beta \in [0,1]$ and define the function $g_i : [0,1] \to \mathbb{R}$ by

$$g_i(\lambda) = h_i(\lambda v_i(\beta)), \quad i \in N^- \cup N^+.$$

Notice that for each $i \in N^+$ the function g_i is concave, while the function g_b is affine. We first show that for each $\lambda \in [0,1]$ and each $i \in N^+$

$$g_b(\lambda) \leq g_i(\lambda). \tag{7.3}$$

The inequality (7.3) is true for $\lambda = 1$ since $g_i(1) = h_i(v_i(\beta)) = \beta$ for each $i \in N^+$. To see that (7.3) holds for $\lambda = 0$ notice that $g_i(0) = h_i(0) = w_i/(w_i - \bar{w})$. Since $w_i \leq w^+ = w_b$ we have $g_b(0) \leq g_i(0)$ for each $i \in N^+$. At last, we prove (7.3) for an arbitrary $\lambda \in [0,1]$. It holds that

$$\begin{aligned} g_b(\lambda) &= g_b[(1-\lambda)0 + \lambda 1] \\ &= (1-\lambda)g_b(0) + \lambda g_b(1) \\ &\leq (1-\lambda)g_i(0) + \lambda g_i(1) \\ &\leq g_i[(1-\lambda)0 + \lambda 1] \\ &= g_i(\lambda), \end{aligned}$$

where the second equality follows because g_b is an affine function and the inequality in the last line holds because g_i is a concave function.

Setting $\lambda = \alpha^+$ and using (7.3) yields

$$f_{1b}(\beta) = g_b(\alpha^+) \leq g_i(\alpha^+) = f_{1i}(\beta), i \in N^+.$$

□

We thus conclude that the tax equilibrium of the society S for any value of the discount factor is also the tax equilibrium of the society T. It follows that the procedurally fair tax rate of the society S equals that of T, implying the following result.

THEOREM 7.8: *The tax rate β^{**} is the procedurally fair tax rate of the society S.*

When the income distribution is sufficiently dispersed, i.e. the case where $w^- = 0$ and w^+ tends to infinity, we reproduce the result we obtained before for societies populated by constant relative risk aversion citizens.

COROLLARY 7.9: *For a sufficiently dispersed income distribution, the procedurally fair tax rate is equal to the probability mass m^- of below average income citizens.*

In conclusion of the section we remark that Assumption 7.3 which clearly drives the results of this section, can be weakened: Rather than requiring that the citizens a and b be present in the society we can only ask that there be the citizens whose characteristics are sufficiently close to those of a and b. More precisely, we can replace Assumption 7.3 by the following assumption.

ASSUMPTION 7.10: There exists a sequence $(t^n)_{n\in\mathbb{N}}$ of types in N with the property that $w_{t_n} \to w^-$ and $u_{t_n}(c) \to c$ for each $c \geq 0$. There exists a sequence $(s_n)_{n\in\mathbb{N}}$ of types in N with the property that $w_{s_n} \to w^+$ and $u_{s_n}(c) \to c$ for each $c \geq 0$.

All results stated in this section continue to be true under this, weaker, version of Assumption 7.3, with the society T defined exactly as before. Only the proof of Lemma 7.7 has to be extended, as follows.

PROOF OF LEMMA 7.7: One shows, exactly as before, that $f_{1b}(\beta) \leq f_{1i}(\beta)$ for each $i \in N^+$. Let the sequence $(s^n)_{n\in\mathbb{N}}$ be as in Assumption 7.10. Since the sequence of functions $(u_{s_n})_{n\in\mathbb{N}}$ converges pointwise to u_b, the sequence $(f_{1s_n})_{n\in\mathbb{N}}$ converges pointwise to f_{1b}. It follows that

$$f_{1b}(\beta) = \inf_{i\in N^+} f_{1i}(\beta), \quad \beta \in [0,1].$$

The result follows. □

The result that emerges from this paper is that under a wide variety of circumstances, m^- appears as the procedurally fair tax rate. This result is in sharp contrast to the literature on fairness that often argues in favor of complete redistribution, implying a tax rate equal to one. The result is also not out of line of what is observed in reality. Mankiw, Weinzierl, and Yagan (2009) report that the average top marginal tax wedge in 2007, which combines the top marginal income tax rate with the rate of value-added tax, is just above 60 percent in OECD countries, and was in fact nearly 80 percent in 1984. Moreover, in this paper we assume a completely inelastic labor supply, and we do therefore not take into account the negative incentive effects that result from taxation. Doing so is likely to reduce the procedurally fair tax rate to values below m^-.

8 Conclusion

We study the implications of procedural fairness on income taxation. Procedural fairness is distinct from fairness and avoids the postulation of an ad hoc social welfare function that has to be optimized. Societies are shown to have unique procedurally fair tax rates. We also provide a characterization of the procedurally fair tax rate, which can be used to demonstrate that the procedurally fair tax rate is equal to the probability mass m^- of below average income citizens in a variety of circumstances.

This paper has confined itself to a rather simple economic environment in order to obtain sharp results. Many extensions of the model are worthwhile to investigate. First and foremost would be to allow for elastic labor supply of citizens, the main channel through which complete redistribution is avoided in the traditional optimal taxation literature. One would expect that also in our set-up the incorporation of elastic labor supply leads to a reduction of the optimal tax rate because of the usual incentive effects.

We have limited ourselves to affine tax schedules, meaning a fixed subsidy to all citizens and constant marginal tax rates. Mirrlees (1971) argued that affine tax schedules are nearly optimal in the context of the traditional optimal taxation literature. Though this view has been challenged many times, Mankiw, Weinzierl, and Yagan (2009) claim that proposals for a flat tax are not inherently unreasonable. It is an open issue whether flat tax schedules are also nearly optimal in our framework.

References

Aumann, R.J., and M. Kurz (1977), "Power and Taxes," *Econometrica, 45,* 1137–1161.

Banks, J., and J. Duggan (2000), "A Bargaining Model of Collective Choice," *American Political Science Review, 94,* 73–88.

Binmore, K. (1987), "Perfect Equilibria in Bargaining Models," in K. Binmore and P. Dasgupta (eds.), *The Economics of Bargaining,* Basil Blackwell, Oxford, UK, pp. 77-105.

Bolton, G.E., J. Brandts, and A. Ockenfels (2005), "Fair Procedures: Evidence from Games Involving Lotteries," *The Economic Journal, 115,* 1054–1076.

Cardona, D., and C. Ponsatí (2007), "Bargaining One-dimensional Social Choices," *Journal of Economic Theory, 137,* 627–651.

Cho, S.J., and J. Duggan (2003), "Uniqueness of Stationary Equilibria in a One-dimensional Model of Bargaining," *Journal of Economic Theory, 113,* 118–130.

Fleurbaey, M., and F. Maniquet (2006), "Fair Income Tax," *Review of Economic Studies, 73,* 55–83.

Fudenberg, D., and J. Tirole (1991), *Game Theory,* MIT Press, Cambridge, Massachusetts.

Herings, P.J.J., and A. Predtetchinski (2010), "One-dimensional Bargaining with Markov Recognition Probabilities," *Journal of Economic Theory, 145,* 189–215.

Imai, H., and H. Salonen (2000), "The Representative Nash Solution for Two-sided Bargaining Problems," *Mathematical Social Sciences, 39,* 349–365.

Kalandrakis, T. (2006), "Regularity of Pure Strategy Equilibrium Points in a Class of Bargaining Games," *Economic Theory, 28,* 309–329.

Mankiw, N.G., M. Weinzierl, and D. Yagan (2009), "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives, 23,* 147–174.

Mirrlees, J.A. (1971), "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies, 38,* 175–208.

Predtetchinski, A. (2007), "One-dimensional Bargaining with Unanimity Rule," METEOR Research Memorandum 07/11, Maastricht University, Maastricht, pp. 1–30.

Roth, A.E. (1989), "Risk Aversion and the Relationship between Nash's Solution and Subgame Perfect Equilibrium of Sequential Bargaining," *Journal of Risk and Uncertainty, 2,* 353–365.

Rubinstein, A. (1982), "Perfect Equilibrium in a Bargaining Model," *Econometrica, 50,* 97–109.