52 Power and Taxes
with M. Kurz

1 Introduction

Perhaps the most fundamental element in the theory of the public sector is the view that the government is an exogenous, benevolent economic agent. The benevolence of the government is often expressed by assuming it to maximize a "social welfare" function of the form $\int u_t(\boldsymbol{x}(t))\mu(dt)$, where the u_t are utility functions, $\boldsymbol{x}$ is an allocation of consumption bundles and μ is a distribution of agent types; in other words, simply the sum of individual utilities. With such a social welfare function Arrow and Kurz [2] were able to derive optimal investment and taxation programs while Mirrlees [13], Sheshinski [19] and others were able to derive optimal tax policies for a population with heterogeneous endowment.

We do not think that this view is without merit. There are perhaps some public issues with regard to which a consensus may be reached, and then such an approach may suffice to explain the behavior of the government. But more often, redistributive effects are a central issue; and then the actions of the government, and in particular its tax policies, can be understood only as an endogenous consequence of the political forces that enable it to maintain power. For this reason one should investigate the connection between tax policies and the political forces that shaped those policies to begin with.

We propose to regard income distribution, taxation, the production of public goods and other actions of the public sector as determined by a political process simultaneously with the economic process of exchange and production. This means that we propose to study an economic-political equilibrium where the power of each individual is reflected both in the political and the economic spheres.

The present is our first paper on this subject,[1] where we formulate the basic structure and motivate it. All of our work in this paper deals with a world in which there is only one commodity ("money"), to be thought of as an aggregate of all real commodities. This enables us to focus on the purely redistributive aspect of government policy. In another paper [4] we treat a more complex and realistic multi-commodity model in which we

This chapter originally appeared in *Econometrica* 45 (1977): 1137– 1161. Reprinted with permission.

1. This research was supported by U.S. National Science Foundation Grant GS-40104 at the Institute for Mathematical Studies in the Social Sciences at Stanford University, and by a grant from the Israel National Council for Research and Development at the Hebrew University of Jerusalem. We are greatly indebted to Kenneth Arrow for an extremely helpful conversation on the subject of this paper; the interpretation of δ_t in terms of "fear of ruin" (Section 6) is an outcome of that conversation.

discuss such matters as the relations between taxes and prices for various goods.

The basic tool that deals with the conflict part of our theory comes from game theory; it is the "Harsanyi–Shapley–Nash Value for Non-Transferable Utility Games" (Shapley [18]). The highlights of our results are as follows:

a. If a democratic power structure (majority vote) is assumed, then an income tax emerges which can be progressive, regressive or neutral.

b. The size of the tax depends upon attitudes toward risking large losses. We shall introduce a new measure for the attitude toward such risks, which we shall call "the Fear of Ruin." This may be contrasted with the Arrow–Pratt [1, 16] measures of risk aversion, which measure attitudes towards small risks only. The connection with risk is somewhat startling, since there is no overt element of risk in the model; however, a closer examination of the situation, which will be undertaken below, shows that considerations of risk do enter naturally.

c. The tax structure involves a personal support level and negative taxation for low income people. (The "support" is what the individual receives from the government in the absence of any income.)

d. The marginal tax rate is always between 50 per cent and 100 per cent.

The treatment in this paper will be on the conceptual level, and will not be entirely rigorous from the mathematical viewpoint. A rigorous treatment, with complete proofs, may be found in [4]. The "heaviest" parts of this paper from the mathematical viewpoint are Sections 4 and 9; readers who are willing to forego the precise definition of the Shapley value, and of the corresponding concept of value allocation, may skip these sections.

2 The Income Redistribution Game

In the model of this paper there is only one commodity, to be thought of as an aggregate of all "real" commodities; for convenience we shall call it "money."[2] There is a set T of agents ("society"), each one of whom has an initial endowment of money ("gross income"). It is assumed that there are many agents, and that each individual one is "small;" that is, remov-

2. We use the term "money" where often economists use the term "income." We wish to keep the ideas behind these terms distinct: in our terminology, the commodity "money" provides the units in which wealth or income are measured. Thus we can talk of "income," "gross income," and "net income" as specific quantities of money received by individuals (or groups) in various circumstances.

ing him would not appreciably affect society. By a political procedure to be specified below, the agents in T will be taxed, and the resulting revenue will be distributed among them; what is left to an agent after that is called his "net income" (it may exceed gross income). We shall be interested only in the relation between gross and net income—i.e., in the total effect of both taxation and redistribution—and not in the two processes separately.

The political procedure is majority rule. This means that any coalition containing more than half the agents can impose taxes and then redistribute them in any way it pleases. In particular, it may divide among its own members the taxes collected from the minority.

It might be objected that in a democracy, one aims at "uniform" tax laws, i.e. laws under which people with the same income are taxed in the same way; this would preclude taxing two people differently just because one is in the ruling coalition while the other is not. But in fact, tax laws are not uniform, as witness the myriads of different rules for different kinds of taxpayers and different kinds of income—rules that are often deliberately slanted to favor various pressure groups.[3] Moreover, it must be recalled that what we are discussing here is not just taxation, but rather the total effect of government activity on incomes. Though theoretically one might require uniform taxation, there can be no such requirement on government spending; and spending can easily be determined by the majority to wipe out any undesired effects of uniform taxation requirements. Finally, we are even willing to concede that total income redistribution may in fact exhibit certain aspects of uniformity, if the concept is interpreted broadly. But we would argue that this is the *result* of democratic forces at work—a compromise between clashing pressure groups—rather than a precondition for democracy. In formal terms, we would expect such uniformity as an *outcome* of the theory; a conclusion, not a hypothesis.

Since the majority can, in principle, tax the minority at 100 per cent and return nothing to it, it would seem at first sight that the total endowment of society becomes available to whomever is in the majority. Thus individual endowments would lose their significance, and we would be led to a completely egalitarian analysis. While this kind of analysis is not without interest, it appears too extreme to be considered relevant to the problem of income distribution in a democracy.

We now appear to be caught in a dilemma. On the one hand, majority rule appears to lead immediately to the extreme of total egalitarianism.

3. No adverse value judgment is intended. Pressure groups are what democracy is all about—they are as essential to healthy politics as competition is to a healthy economy.

On the other hand, if one ignores politics and does not allow the majority to redistribute income, one is simply left with the initial income distribution. How, then, can one account politically for the type of taxation scheme that is observed, in which some redistribution takes place but net income still remains related to gross income?

To answer this question, we will make an assumption that we consider a basic ingredient of a democratic society, namely that

every agent can, if he wishes, destroy part or all of his endowment.

It goes without saying that the part that is destroyed cannot be taxed. If one thinks of one's endowment as labor, then the above means that there is *no forced labor*: an individual may, if he wishes, "destroy" his labor, by simply working less (or not at all).[4]

It may not be immediately clear why this assumption changes anything—after all, who would want to destroy his endowment—what good would it do to anybody? The answer is that it gives the minority very considerable *threat power*—power that is vital in determining taxes. There are numerous cases in history where farmers' revolts against the tax collector were associated with the destruction of crops; this was recently demonstrated by the French farmers of Normandy who destroyed their crops on the roads against President De Gaulle. Also, a strike involves the destruction of endowment (in the form of labor services) in the face of what are considered unfavorable terms of trade or excessive taxation. Though the minority certainly cannot guarantee to itself any part of its endowment, nevertheless it can say to the majority "it will be neither mine nor thine."[5] This is a powerful threat, which can force the majority to compromise. And it is this compromise that underlies the delicate balance between individual rights and Society's needs inherent in all tax schemes.

There is another element that enters the description of our model. Every agent is assumed to have a von Neumann–Morgenstern (N–M) utility [20] for money. The reader may ask, of what possible relevance can N–M utility functions—which really measure attitudes toward risk—be in a situation that contains no overt elements of risk? The answer is that on the contrary, we are dealing with a situation that is replete with risky elements. The bargaining associated with entry into (or exclusion from) a ruling coalition is a risky business, as are the threats of the minority and of the majority once coalitions have been formed. One's

4. In this interpretation no positive utility is assigned to leisure. See the discussion in Section 10, Subsection b.

5. I Kings 3, 26.

attitude toward these risks is a decisive factor in how well one can do in the bargaining, even though in the end no random mechanism is used to determine the payoff, i.e. no risks are taken by anyone.

To sum up, the *Income Redistribution Game* is played as follow: each agent starts out with an endowment and a utility function. Redistribution decisions are made by majority vote, but each agent has the right to destroy some or all of his endowment.

Our results turn out to be insensitive to the exact strategic description of the game (e.g. whether destruction decisions are made before, after, or simultaneously with announcement of tax laws) and therefore we will leave these matters unspecified.

3 The Formal Description of Society

Society—the set of "agents" or "players"—will be denoted T. We will model our assumption that society consists of many individually small agents by taking T to be a *continuum*[6] (like a line or a region) rather than a finite set. In such a model it is best to think of an agent as an infinitesimal subset dt of T; however, we will usually still find it convenient to label agents by points t in T, where the label t is to be thought of as a "typical" point in the infinitesimal set[7] dt.

Next, we require some formal way of measuring the size of coalitions (i.e., sets of agents), in order to be able to say when they are in the majority. If T were finite one could do this simply by counting. Since T is a continuum, counting will not do, and one must specify the measure of size exogenously. We therefore assume given a nonnegative measure[8] μ on the coalitions, with $\mu(T) = 1$ (the *population measure*). Intuitively, one should think of $\mu(S)$ as the proportion of traders in S; thus S is in the majority if and only if $\mu(S) > 1/2$. Note that the measure $\mu(dt)$ of a single agent can be thought of as the reciprocal of the number of people in society.

Finally, for each t in T, there is given a nonnegative number $\boldsymbol{e}(t)$ (t's *endowment*) and a function u_t on the nonnegative numbers (t's *utility*). The endowment function $\boldsymbol{e}$ is to be thought of as a density function, like

6. Compare Aumann and Shapley [5], especially Section 29, or Hildenbrand [10]. Technically T is taken as a measurable space, isomorphic to the unit interval with the Borel subsets. Only measurable sets are to be thought of as coalitions.

7. One may think of T as an interval that is cut up into a large number of small subintervals dt, each of which is labelled by a point t in it.

8. A *measure* is a function μ on the coalitions such that $\mu(S \cup W) = \mu(S) + \mu(W)$ when S and W are disjoint, and more generally, $\mu(\bigcup_{i=1}^{\infty} S_i) = \sum_{i=1}^{\infty} \mu(S_i)$ when the S_i are disjoint.

the density functions that occur in probability theory; the actual endowment of an agent t is $\boldsymbol{e}(t)\mu(dt)$, and the total endowment of a coalition S is $\int_S \boldsymbol{e}(t)\mu(dt)$. If $\boldsymbol{x}$ is a function on T, denote $\int_S \boldsymbol{x}(t)\mu(dt)$ simply by $\int_S \boldsymbol{x}$, and $\int_T \boldsymbol{x}(t)\mu(dt)$ by $\int \boldsymbol{x}$. Thus the total endowment of a coalition S is $\int_S \boldsymbol{e}$.

We shall require some assumptions. The first is:

The population measure μ *is nonatomic.* (3.1)

Nonatomicity means that T can be cut up into coalitions all of which have μ-measure as small as we like; intuitively, it means that the agents are individually negligible. We shall also assume:

The u_t *are increasing, concave, continuously differentiable at positive values of the argument, and continuous at* 0. (3.2)

$$\infty > \int \boldsymbol{e} > 0. \tag{3.3}$$

The u_t *are uniformly bounded,*[9] *and* $u_t(1)$ *is uniformly positive.*[10] (3.4)

$$u_t(0) = 0. \tag{3.5}$$

Assumptions (3.1) and (3.2) are substantive, and without them we could not prove our result. The first half of (3.3)—that $\infty > \int \boldsymbol{e}$—merely says that $\boldsymbol{e}$ is integrable, i.e. that the total wealth of society is finite, whereas the second half—that $\int \boldsymbol{e} > 0$—merely says that some significant set of agents has a positive endowment, without which the whole situation becomes trivial. Assumption (3.4) is also of a technical nature, and it might be possible to dispense with it or at least weaken it. Assumption (3.5) is, of course, merely a normalization.

There are also some very technical measurability assumptions, which we will not spell out here. The interested reader is referred to [4].

As usual, an *allocation* is a nonnegative function $\boldsymbol{x}$ on T such that $\int \boldsymbol{x} = \int \boldsymbol{e}$; here, too, we must view $\boldsymbol{x}$ as a density function. If $\boldsymbol{x}$ is a nonnegative function on T, we shall abuse our notation slightly by writing $u(\boldsymbol{x})$ for the function on T whose value at t is $u_t(\boldsymbol{x}(t))$; the notation $u'(\boldsymbol{x})$ is to be interpreted analogously, u'_t being the first derivative of u_t. Since all individual quantities are scaled down to infinitesimal size, it is useful to think even of the utilities as densities; i.e., to think of t's utility for an allocation $\boldsymbol{x}$ as $u_t(\boldsymbol{x}(t))\mu(dt)$.

9. $\sup_{x,t} u_t(\boldsymbol{x}) < \infty$.

10. $\inf_t u_t(1) > 0$. This implies that $\inf_t u_t(x) > 0$ for all $x > 0$.

4 The Solution Concept

The game described in Section 2 is "played" on two levels. First, there is maneuvering to determine which players will be in the majority and minority coalitions, respectively; and then, there is bargaining between the two coalitions that actually form, involving threats by both sides. Our aim is to find an outcome that avoids conflict and assigns to each player a payoff reflecting his opportunities on both levels of play; in short, a reasonable compromise. It is our view that the notion most suitable for this purpose is the concept of value introduced by Shapley [17], as adapted to variable threat games by Harsanyi [9], to nontransferable utility (NTU) games by Shapley [18], and to games with a continuum of players by Kannai [11] and Aumann and Shapley [5]. The underlying ideas date back to the Nash variable threat model [15]. This history explains why we will refer to the solution concept adopted here as the Harsanyi–Shapley–Nash NTU value.

To explain this solution concept, we start with finite games. Recall that a *finite coalitional game* (or simply *game*) consists of a finite set T (the "players") together with a function v that associates with each subset S of T ("coalition") a real number $v(S)$ (the "worth" of S), such that $v(\emptyset) = 0$. A *payoff vector* is a measure ξ on the coalitions; intuitively, $\xi(S)$ represents the sum of the payoffs to all members of S. Since T is finite, ξ is determined by its values $\xi(\{t\})$ on one-player sets, so that it may be thought of as the vector of payoffs to the individual players.

A *value* [17] is an operator that associates with each game v a payoff vector ϕv satisfying certain plausible axioms. Here we will quote only the *efficiency* axiom, according to which

$$(\phi v)(T) = v(T), \tag{4.1}$$

and the *additivity* axiom, according to which

$$\phi(v + w) = \phi v + \phi w. \tag{4.2}$$

Shapley [17] also showed that the value is given by the formula

$$(\phi v)(\{t\}) = E(v(S_t^{\mathscr{R}} \cup \{t\}) - v(S_t^{\mathscr{R}})), \tag{4.3}$$

where $S_t^{\mathscr{R}}$ is the set of players preceding t in a random order $\mathscr{R}$ on the set of all players, and E is the expectation operator when all orders on T are assigned equal probability. Formula (4.3) means that the value of a player is the expectation of his contribution to the worth of the players preceding him in a random order of all players.

The concepts of coalitional game and value discussed above have been extended from a finite player set to a continuum of players; see [5, 11]. As in the finite case, the value associates to a coalitional game v a payoff vector ϕv, where by "payoff vector" we mean a measure on coalitions.

We turn now to the problem of defining a coalitional game corresponding to the income redistribution game described in Section 2. This means that we need to specify the "worth" $v(S)$ of each coalition S. In attempting to do that we are faced with two basic difficulties: First, the utilities of the players are not "comparable," so that it is meaningless to speak of one players getting, say, twice as much utility as another; on the other hand, any measure of the worth of a coalition must obviously involve some kind of aggregation of the utilities of the individuals, and it is difficult to see how such an aggregation can be carried out in the absence of comparability. The second difficulty is that even if we could somehow aggregate utilities to determine how much each outcome is worth to each coalition, we still would not know how to determine the worth of a coalition. This is because even after coalitional lines are drawn, the majority and the minority are faced with a fairly complex strategic situation in which threats, counterthreats, and compromises play a crucial role. It is not at all clear a priori how these strategic considerations would interact and to what outcome they would lead.

Recall that since u_t is an N–M utility for t, it follows that for any positive constant λ, λu_t is also an N–M utility. In formal terms, the problem of comparability is that the constants λ can be chosen entirely arbitrarily, with no restriction that they must be the same for different players. Solving this problem means finding a criterion of comparability—i.e., somehow fixing a "weight" $\lambda = \boldsymbol{\lambda}(t)$ for each agent t. If that were done, then one could define the aggregate utility of a coalition S for an allocation $\boldsymbol{x}$ by the expression[11] $\int_S \boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t))\mu(dt) = \int_S \boldsymbol{\lambda}u(\boldsymbol{x})$.

Let us for the moment postpone discussing the choice of weights λ, and proceed at once to the second question, namely that of determining the worth of a coalition in view of the strategic possibilities open to it and its complement. Suppose, then, that the function λ is given; we wish to determine the worth $v_{\boldsymbol{\lambda}}(S)$ of a coalition S. Consider first the case $S = T$. This is easier because the assumption that the coalition T of all agents has "formed" means that all agents are cooperating; hence it is only a question of maximizing total aggregate utility $\int \boldsymbol{\lambda}u(\boldsymbol{x})$ subject to $\int \boldsymbol{x} = \int \boldsymbol{e}$, and there is no question of threats, counterthreats, and compromises. To achieve this maximum, the players will presumably reallocate the initial endowment between them in some way; the resulting distribution of net

11. This notion of "aggregate utility" is of course not without conceptual difficulties; but its intuitive meaning in connection with the NTU value has been discussed by Shapley [18] and Aumann [3], and we will not repeat the discussion here.

income will be an allocation, i.e. $\int \boldsymbol{x} = \int \boldsymbol{e}$, since total net income must equal total gross income (remember that we are only discussing redistribution of income). The aggregate utility of T will then be $\int \boldsymbol{\lambda} u(\boldsymbol{x})$. Since the coalition T will act to maximize this aggregate utility, we may conclude that[12]

$$v_\lambda(T) = \max\left\{\int \boldsymbol{\lambda} u(\boldsymbol{x})\colon \int \boldsymbol{x} = \int \boldsymbol{e}\right\}. \tag{4.4}$$

Suppose next that $S \neq T$. We will think of $v_\lambda(S)$ as being the aggregate utility of S if it "forms" and bargains as a unit with the agents outside of S. In analyzing this situation, we use a simplified version of the model of Nash [15]. Suppose both S and its complement $T\setminus S$ commit themselves to carrying out certain threats—let's call them σ and τ respectively—if an accomodation between them is not reached. Let's say that carrying out these threats will yield to S and $T\setminus S$ aggregate utilities of, say, $f = f_\lambda(\sigma, \tau)$ and $g = g_\lambda(\sigma, \tau)$ respectively. Thus after the threats are made, the surplus utility over which they are bargaining (and that somehow has to be split between them) is $v_\lambda(T) - f - g$, where $v_\lambda(T)$ is given by (4.4). Under these circumstances, symmetry conditions (cf. Nash [14]) indicate a compromise in which this surplus will be evenly divided between them, so that the resulting aggregate utility to S should be

$$f + \frac{v_\lambda(T) - f - g}{2} = \tfrac{1}{2}[v_\lambda(T) + (f - g)], \tag{4.5}$$

and similarly, to $T\setminus S$ it should be

$$\tfrac{1}{2}[v_\lambda(T) - (f - g)]. \tag{4.6}$$

Of course each of the sides will try to choose its threat in a way that is most advantageous to its final payoff. Hence, if we set

$$H = H_\lambda^S(\sigma, \tau) = f - g = f_\lambda(\sigma, \tau) - g_\lambda(\sigma, \tau),$$

then we may conclude from (4.5) and (4.6) that S will wish to maximize, and $T\setminus S$ to minimize, the expression H. By the minimax theorem[13] for 2-person 0-sum games, there is a number $w = w_\lambda(S)$ such that S can guarantee that H will be at least w, and $T\setminus S$ can guarantee that it will be at most w. Presumably S and $T\setminus S$, if formed would act to "cash

12. Fine points such as questions of existence of the maximum will be steadfastly ignored in this paper.

13. We remind the reader that we are ignoring "fine points" such as verifying that the conditions for the minimax theorem hold.

in" on these guarantees, so that one could then expect an outcome that yields to S and $T\backslash S$ aggregate utilities of $1/2(v_\lambda(T)+w_\lambda(S))$ and $1/2(u_\lambda(T)-w_\lambda(S))$ respectively. Thus given the function λ, we may define the worth $v_\lambda(S)$ by

$$v_\lambda(S)=\tfrac{1}{2}(v_\lambda(T)+w_\lambda(S)). \tag{4.7}$$

We are interested in the value ϕv_λ of the game v_λ.

But there is a difficulty here. To see it note first the value ϕv_λ calls for a "redistribution of utilities," since the worths $v_\lambda(S)$ defined in (4.4) are in units of "aggregate utility." In fact, we cannot redistribute utility, but only income; thus it is possible that the value ϕv_λ is not attainable—i.e., there is no redistribution of income (or allocation) that yields each agent the utility that he is assigned by the payoff vector ϕv_λ. In fact, we might remark parenthetically that in general there is not even any allocation that will yield S and $T\backslash S$ the amounts (4.5) and (4.6); that is because (4.5) and (4.6) involve an even split of surplus aggregate *utility*; and in our model, utility, though one may assume it interpersonally comparable, is certainly not transferable—only money can be transferred. Thus the compromises implicit in (4.5) and (4.6), as well as the larger compromise implicit in the value ϕv_λ, may simply be infeasible.

And as if that were not sufficiently worrisome, we must remind the reader that we have not yet resolved the problem of determining the weights λ.

Fortunately, these two difficulties cancel each other out. We will show that there is a unique[14] function λ such that ϕv_λ is feasible; thus the feasibility problem and the problem of determining λ solve each other. Formally, we define a *value allocation* to be an allocation $\boldsymbol{x}$ such that there exists a λ for which

$$(\phi v_\lambda)(dt)=\lambda(t)u_t(\boldsymbol{x}(t))\mu(dt) \tag{4.8}$$

for all agents t (or equivalently, $(\phi v_\lambda)(S)=\int_S \lambda u(\boldsymbol{x})$ for all S). This means that the value assigns to each agent precisely the utility he receives under the allocation $\boldsymbol{x}$. Thus the value is feasible without any "transfers of utility;" the only transfers are those of money.

The Harsanyi–Shapley–Nash value just defined, and the corresponding concept of value allocation, have been thoroughly discussed elsewhere [3, 15, 18], and we will not enter into another such discussion here. A few words about the concept may, however, be in place. Integrating (4.8) and

14. This is true only in the income redistribution game described here; in general one will not get uniqueness of the λ in an NTU game, though the set of appropriate λ may be expected to be in some sense "small."

using the efficiency axiom (4.1), we find

$$v_\lambda(T) = (\phi v_\lambda)(T) = \int \lambda u(\boldsymbol{x}).$$

From this and (4.4) it follows that the maximum that defines $v_\lambda(T)$ is achieved at the value allocation $\boldsymbol{x}$; hence the marginal utilities $\lambda(t)u_t'(\boldsymbol{x}(t))$ must be equal for all agents t with positive income. This means that although utility itself is not transferable, if one scales the utilities u_t by multiplying them by $\lambda(t)$, then in the neighborhood of a value allocation, utility is "locally transferable;" i.e., small amounts of utility can effectively be transferred by transferring money.

Of course for any allocation $\boldsymbol{x}$ one can find λ's with this property—it is only necessary to choose $\lambda(t) = 1/u_t'(\boldsymbol{x}(t))$. But when one then proceeds to calculate ϕv_λ, one will in general find that it is infeasible. Thus the value allocation is uniquely determined by requiring utility comparisons to be made in such a way so that utility is locally transferable, *and* that the allocation be a reasonable compromise, from the point of view of the Shapley value.

5 The Main Theorem: Statement and Preliminary Discussion

MAIN THEOREM *The income redistribution game has a unique value allocation $\boldsymbol{x}$. This allocation satisfies*

$$\frac{u_t(\boldsymbol{x}(t))}{u_t'(\boldsymbol{x}(t))} + \boldsymbol{x}(t) = c + \boldsymbol{e}(t) \tag{5.1}$$

for all t, where c is a positive constant.

Note that

$$c = \int \frac{u(\boldsymbol{x})}{u'(\boldsymbol{x})}; \tag{5.2}$$

this follows from integrating (5.1) and using $\int \boldsymbol{x} = \int \boldsymbol{e}$.

We will demonstrate the Main Theorem informally in Section 9. A formal proof is given in [4].

Let us explore some implications of the main theorem. First, we have

IMPLICATION 5.3 *For agents with the same utility function, net income is an increasing function of gross income $\boldsymbol{e}(t)$.*

This follows from the fact that by Assumption (3.2), the left side of (5.1) is an increasing function of $\boldsymbol{x}(t)$.

Next, note that

$$t\text{'s income tax}^{15} = \boldsymbol{e}(t) - \boldsymbol{x}(t). \tag{5.4}$$

Hence, the *marginal income tax rate* $\boldsymbol{M}(t)$ is the derivative of $e - x$ with respect to e at $e = \boldsymbol{e}(t)$, when x is implicitly defined as a function of e by the equation

$$\frac{u_t(x)}{u_t'(x)} + x = c + e.$$

Implicit differentiation of this equation yields

IMPLICATION 5.5 *Assume that u_t is twice continuously differentiable. Then*

$$\boldsymbol{M}(t) = 1 - \frac{1}{2 + (-u_t'' u_t / u_t'^2)}, \tag{5.6}$$

where u_t and its derivatives are calculated at $\boldsymbol{x}(t)$.

In particular,

$$\tfrac{1}{2} \leqq \boldsymbol{M}(t) < 1, \tag{5.7}$$

i.e., the marginal rate is always at least 50 per cent! In Section 10 we return to this matter of the surprisingly high marginal rate.

Let us now interpret the main theorem in terms of the total rather than the marginal tax. From (5.1) and (5.4) we obtain

$$t\text{'s income tax} = \frac{u_t(\boldsymbol{x}(t))}{u_t'(\boldsymbol{x}(t))} - c. \tag{5.8}$$

Since $\boldsymbol{x}(t)$ is the allocation of income to agent t and since $u_t'(\boldsymbol{x}(t))$ is his marginal utility of income, $\boldsymbol{\lambda}(t) = 1/u_t'(\boldsymbol{x}(t))$ is the *money price* of a unit of t's utility. Thus in the final compromise, agent t receives utility with money valuation $\boldsymbol{V}(t)$ defined by

$$\boldsymbol{V}(t) = \boldsymbol{\lambda}(t) u_t(\boldsymbol{x}(t)) = \frac{u_t(\boldsymbol{x}(t))}{u_t'(\boldsymbol{x}(t))}. \tag{5.9}$$

Combining (5.1), (5.2), and (5.9), we find

IMPLICATION 5.10 *t's income tax $= \boldsymbol{V}(t) - \overline{V}$, where $\overline{V}$ is the average*[16] *of the $\boldsymbol{V}(t)$.*

15. It should be recalled that we are really discussing the net effect of taxation *and* redistribution; use of the word "tax" to describe the combination of both processes is merely a matter of convenience. Note also that the decrement in t's income is actually $(\boldsymbol{e}(t) - \boldsymbol{x}(t))\mu(dt)$; thus $\boldsymbol{e}(t) - \boldsymbol{x}(t)$ is really "tax density" rather than "tax."

16. $\overline{V}$ is of course the same as the constant c of (5.1) and (5.2); we use the notation $\overline{V}$ only to emphasize the interpretation of this constant as a mean of the $\boldsymbol{V}(t)$.

This means that a person's tax is the surplus of the dollar value of his utility over the average of the dollar values of everybody's utilities. Thus we get a tax structure which is negative on the lower part of the scale, i.e. provides for positive public transfer to low income agents. In fact if we denote the elasticity of utility by

$$\boldsymbol{\eta}(t) = \frac{u_t'(\boldsymbol{x}(t))}{u_t(\boldsymbol{x}(t))}\boldsymbol{x}(t),$$

then (5.1) can be solved to read

$$\boldsymbol{x}(t) = \frac{\boldsymbol{\eta}(t)}{1+\boldsymbol{\eta}(t)}(c+\boldsymbol{e}(t)).$$

If agent t has no income, i.e. $\boldsymbol{e}(t) = 0$, then this yields a social "support" (or negative tax) to agent t of size $[\boldsymbol{\eta}(t)/(1+\boldsymbol{\eta}(t))]c$. As his income increases the agent pays a positive marginal income tax $\boldsymbol{M}(t)$ until the support is exhausted at which time the agent begins to pay a positive amount of total taxes.

6 The Fear of Ruin

In the discussion above (at (5.8)), we saw that

$$t\text{'s income tax} = \frac{u_t(\boldsymbol{x}(t))}{u_t'(\boldsymbol{x}(t))} - c,$$

where c may be viewed as a constant tax credit. Let us now interpret the term $u_t(\boldsymbol{x}(t))/u_t'(\boldsymbol{x}(t))$, in terms of behavior under uncertainty.

For simplicity, set $x = \boldsymbol{x}(t), u = u_t(x), u' = u_t'(x)$. The expression u/u' will be called t's *fear of ruin* at x. To understand why, let us consider its reciprocal u'/u, which will be called t's *boldness* at x. Suppose that t is considering a bet in which he risks his entire fortune x against a possible gain of a small amount h. The probability q of ruin would have to be very small in order for him to be indifferent between such a bet and retaining his current fortune. Moreover, the more unwilling he is to risk ruin, the smaller q will be. Thus q is an *inverse* measure of t's aversion to risking ruin, and a direct measure of boldness; obviously q tends to 0 as the potential winnings h shrink. We assert that the boldness is the *probability of ruin per dollar of potential winnings* for small potential winnings, i.e., it is the limit of q/h as $h \to 0$. To see this, note that for indifference we must have

$$u(x) = (1-q)u(x+h) + qu(0) = (1-q)u(x+h).$$

Hence

$$\frac{q}{h} = \frac{(u(x+h) - u(x))/h}{u(x+h)}$$

and as $h \to 0$, this tends to u'/u.

We conclude that *the tax equals the fear of ruin at the net income, less a constant tax credit*. Thus the more fearful a person, the higher he may expect his tax to be.

Next, set $u'' = u_t''(x)$, and let us examine the term

$$\delta_t = -uu''/u'^2 \tag{6.1}$$

appearing in the expression (5.6) for the marginal tax rate. We have

$$\delta_t = \frac{-u''/u'}{u'/u}.$$

The denominator of the right side is the boldness; its numerator is the measure of *absolute risk aversion* (of t at x), as defined by Arrow [1] and Pratt [16]. To interpret this concept, let us suppose that t considers an even-money bet in a small amount. Since he is risk averse (i.e. $u'' < 0$), the probability p for success would have to be greater than 1/2 in order for him to be indifferent between this bet and simply retaining his current fortune x. The probability premium $p - 1/2$ is a measure of his aversion to risk at x; because u is differentiable, it tends to 0 as the size of the bet shrinks. The measure of absolute risk aversion is the probability premium per dollar of bet size for small bets, i.e., it is the limit of $(p - (1/2))/h$ as $h \to 0$, where h is the size of the bet.

Conceptually, there are two components that enter into the boldness coefficient. One is t's attitude toward risking his fortune; the other is his attitude toward winnings. We know that he is risk averse, i.e., that even at even money he requires a probability premium for entering into a risk. Obviously this aversion is a factor in determining the probability q that entered into the calculation of boldness. If we wish to measure his attitude "purely" toward risking his fortune, we must somehow cancel out the component that measures his aversion to small risks. This motivates us to define the *pure boldness* as the ratio between boldness and absolute risk aversion, i.e. as $-(u'/u)/(u''/u')$. The *pure fear of ruin* is defined as the reciprocal of the pure boldness, i.e. the δ_t in formula (6.1).

The absolute boldness has the same dimensions as the absolute risk aversion, namely 1/dollar. Therefore the pure boldness and pure fear of ruin are both dimensionless, i.e., invariant under changes in the unit of utility and/or the unit of money.

Returning now to the taxation picture, we see from (5.6) that *the marginal tax rate depends only on the pure fear of ruin δ_t and is directly related to it*: the higher the pure fear of ruin, the higher is the marginal tax rate.

The concept of fear of ruin introduced here is of interest also in an entirely different context, namely that of the Nash Bargaining Problem [14]. Consider a case in which two players are bargaining over the division of a fixed amount R of money, and suppose that their utility functions are u_1 and u_2 respectively, where $u_i(x)$ represents the utility of i for receiving the amount x out of this bargaining process. Assume that the u_i are concave and differentiable and $u_i(0) = 0$. Then the Nash solution calls for maximizing $u_1(x_1)u_2(x_2)$ subject to $x_1 + x_2 = R$. This maximum is attained at the unique point (x_1^*, x_2^*) for which $x_1^* + x_2^* = R$ and

$$\frac{u_1'(x_1^*)}{u_1(x_1^*)} = \frac{u_2'(x_2^*)}{u_2(x_2^*)}. \tag{6.2}$$

We thus find that the Nash solution calls for that compromise which makes the two players *equally fearful of ruin*, where ruin is here taken to mean disagreement. This provides an alternative interpretation of the Nash solution, which has sometimes been ignored by economists because of its strongly cardinal nature. (For another interpretation, see Harsanyi [8].)

7 Illustrations

In this section we will assume that the endowment function $\boldsymbol{e}(t)$ is bounded. From (5.1) it follows that $\boldsymbol{x}(t) \leqslant c + \boldsymbol{e}(t)$. Since $\boldsymbol{e}$ is bounded, it follows than in any particular example net income of all individuals must lie in a fixed finite interval, which we shall call the "relevant range." The behavior of the utility functions outside of this interval is irrelevant—the results do not change if we change the utility functions outside of the relevant range. In the illustrations of this section, we will describe the utilities only in the relevant range; outside of that, they can be chosen arbitrarily as long as they remain bounded and sufficiently differentiable. For example, when we refer to "linear utilities," we mean "utilities that are linear in the relevant range;" the utilities cannot, of course, be linear on the entire real line, since that would violate the boundedness condition.

Example 7.1 Linear Utilities. One's first impression is that the threat possibilities make the rich (i.e. high endowment) players extremely powerful: not only do they have a higher endowment, but also more to threaten with. It might even be thought that they would end up richer than before. But this is not the case. From the linearity of the u_t one obtains $(u_t(x)/u_t'(x)) + x = 2x$, and hence $\boldsymbol{x} = (1/2)(c + \boldsymbol{e})$. Integrating, one obtains $c = e(T)$; thus

$$\boldsymbol{x}(t)\mu(dt) = \tfrac{1}{2}\left(\int \boldsymbol{e}\right)\mu(dt) + \tfrac{1}{2}\boldsymbol{e}(t)\mu(dt). \tag{7.2}$$

The expression $\left(\int \boldsymbol{e}\right)\mu(dt)$ represents t's "equal share" of Society's endowment, i.e., what he would get if Society's endowment $\left(\int \boldsymbol{e}\right)$ were divided equally among all agents. Of course, $\boldsymbol{e}(t)\mu(dt)$ and $\boldsymbol{x}(t)\mu(dt)$ are t's gross and net income respectively. Thus (7.2) says that the net income distribution is a 50–50 compromise between the gross income distribution and an entirely egalitarian distribution—a far cry from our above guess that the rich will get richer! What is happening is that by compromising after a threat to destroy its endowment, the minority can hold on to half of its endowment, while the other half goes to the majority. Thus the term $(1/2)\boldsymbol{e}(t)\mu(dt)$ represents that part of one's income that is not taxed; whereas the term $(1/2)\left(\int \boldsymbol{e}\right)\mu(dt)$ represents that part that results from the fact that one has a priori a 50–50 chance of being in the ruling coalition.

In tax terms, we have a tax rate of 50 per cent, with an exemption in the amount of the equal share; this results in a support in the amount of half the equal share, since the exemption can lead to negative taxable income. Alternatively, one could say that everybody gets taxed at a straight 50 per cent (with no support), and that the resulting revenue is redistributed *equally* among the entire population.

Example 7.3 $u_t(x) = x^\alpha$, $0 < \alpha < 1$. Let us try again to guess the result beforehand. Here we have decreasing marginal utility; the less a person has, the greater his marginal utility of income. This kind of situation usually favors the rich, since they are less prone to threats. For example, suppose two people with utility function $\sqrt{x}$ who are worth \$0 and \$10,000 respectively must agree on how to share an additional \$10,000, or else lose the additional amount entirely. Then the Harsanyi–Shapley–Nash value (which in this case boils down to the bargaining solution of Nash [14]) dictates that the richer man will receive approximately \$6,404, while the poorer one will receive only \$3,596 (see (6.2)). Thus, it seems safe to guess that in Example 7.3 the rich will be relatively better off than in Example 7.1.

Unfortunately, we are wrong again. Here

$$\frac{u_t(x)}{u_t'(x)} + x = \frac{\alpha + 1}{\alpha}\boldsymbol{x},$$

and we find

$$\boldsymbol{x}(t) = \frac{1}{\alpha+1}\left(\int \boldsymbol{e}\right) + \frac{\alpha}{\alpha+1}\boldsymbol{e}(t).$$

Again, we get a compromise between the endowment and the equal share, but in a more egalitarian ratio (when $\alpha = 1/2$ it is 2/3 to 1/3). The smaller α—i.e., the more intensely the poor suffer—the more egalitarian the outcome: an effect exactly the opposite of what we had thought!

The explanation is simple. A rich man who finds himself in the minority is subject to ruin (i.e. 0 utility) just as much as a poor man; and therefore in the minority, he is as prone to threats as the poor man. Thus the increased fear of ruin (x/α as compared with x in the linear case) works to the advantage not of the rich man, but of the majority. Since everybody has an equal chance of being in the majority, this tips the scale away from the initial endowment and towards the egalitarian outcome.

In tax terms, we get a tax rate of $1/(\alpha+1)$ (which may be anywhere between 1/2 and 1), an exemption in the amount of the equal share, and a possibility of negative tax. Alternatively, one could say that everybody gets taxed at the rate $1/(\alpha+1)$, with no exemption, and that the resulting revenue is distributed equally among the entire population.

Example 7.4 $u_t(x) = x^{\boldsymbol{\alpha}(t)}$, $0 < \boldsymbol{\alpha}(t) \leqq 1$. Here the fear of ruin varies from agent to agent. One would expect that the more fearful agent—the one with lower $\boldsymbol{\alpha}(t)$—is penalized in the tax structure, since he is more prone to threats. This is indeed the case. We have

$$\frac{u_t(x)}{u_t'(x)} + x = \frac{\boldsymbol{\alpha}(t)+1}{\boldsymbol{\alpha}(t)}x$$

and hence

$$\boldsymbol{x}(t) = \frac{\boldsymbol{\alpha}(t)}{\boldsymbol{\alpha}(t)+1}\left[\left(\int \frac{\boldsymbol{e}}{\boldsymbol{\alpha}+1} \Big/ \int \frac{\boldsymbol{\alpha}}{\boldsymbol{\alpha}+1}\right) + \boldsymbol{e}(t)\right].$$

The tax is given by

$$\boldsymbol{e}(t) - \boldsymbol{x}(t) = \frac{1}{\boldsymbol{\alpha}(t)+1}\left[\boldsymbol{e}(t) - \boldsymbol{\alpha}(t)\left(\int \frac{\boldsymbol{e}}{\boldsymbol{\alpha}+1} \Big/ \int \frac{\boldsymbol{\alpha}}{\boldsymbol{\alpha}+1}\right)\right].$$

This means that the agent who is more fearful of ruin has both a higher tax rate and a lower support.

Example 7.5 $u_t(x) = \log(1+x)$. From (5.6) we get a marginal tax rate of $1 - 1/(2 + \log(1 + \boldsymbol{x}(t)))$, which rises with $\boldsymbol{x}(t)$, and hence also with $\boldsymbol{e}(t)$. Thus in the case of this very classical utility function, the tax is progressive.

Example 7.6 $u_t(x) = x + \sqrt{x}$. We have

$$\frac{-u_t'' u_t}{(u_t')^2} = \frac{1}{2(1+2\sqrt{x})} + \frac{1}{2(1+2\sqrt{x})^2},$$

which decreases as x increases. Hence by (5.6) the marginal tax rate decreases as $\boldsymbol{x}(t)$ increases, and hence also when $\boldsymbol{e}(t)$ does. The tax is thus regressive; the rate goes from near 2/3 when $\boldsymbol{x}(t)$ is small, to near 1/2 when it is large. What is happening is that when x is close to 0, u_t behaves very much like $\sqrt{x}$, and as it moves up, it behaves more and more like x. The results then follow the pattern of Examples 7.1 and 7.3.

In the next section we will see that there are examples in which the tax is progressive, neutral, and regressive at different points in the range of income.

8 Further Discussion of the Marginal Tax Rate

We have seen (5.7) that the marginal rate is always between 1/2 and 1. In Section 7 we saw that the tax may be progressive, regressive, or neutral, i.e. that the marginal rate may be increasing, decreasing, or constant as a function of wealth. This section is devoted to the question of just what we can say about the behavior of the marginal rate, i.e. to characterizing those functions of wealth that can appear as marginal tax rates.

Call a function u *admissible* if it is bounded, concave, increasing, twice continuously differentiable at positive values of the argument, continuous at 0 and satisfies $u(0) = 0$. Let u be admissible, and for $x > 0$, let

$$m(x) = 1 - \frac{1}{2 + (-u''(x)u(x)/u'(x)^2)}; \tag{8.1}$$

the marginal tax rate of an individual with utility function u and net income x is precisely $m(x)$. Clearly

$$1/2 \leqq m(x) < 1. \tag{8.2}$$

We can also say something about the marginal tax rates of the very rich; we have

$$\limsup_{x\to\infty} m(x) = 1. \tag{8.3}$$

Indeed, because of the concavity of u,

$$(x/2)u'(x) \leqq u(x) - u(x/2).$$

Because u is bounded, the right side approaches 0, and hence $\lim_{x\to\infty} xu' = 0$. Arrow [1] has proved that when utility is bounded, $\lim_{x\to\infty} \inf(xu''/u') \geqq 1$. Hence

$$\lim\,\inf(u''u/u'^2) = \lim\,\inf \frac{xu''/u'}{xu'/u} = \frac{\lim\,\inf(xu''/u')}{\lim\, xu'/\lim\, u} = \infty.$$

Hence $\lim\,\sup\, m(x) = 1$, as claimed.

Formula (8.3) says that for arbitrarily large net incomes, the marginal tax rate comes close to 1. On the other hand, it cannot come *too* close too often; letting R^+ be the nonnegative part of the real axis, we have

$1/(1-m)$ *is integrable over any interval in* R^+. (8.4)

Indeed, when the interval has positive end points, this follows from

$$\frac{1}{(1-m)} = 2 - (u''u/u'^2) = (x + (u/u'))';$$

when the left end point is 0, it follows from a limiting argument using $u/u' \to 0$ as $x \to 0$.

The question now arises whether anything can be said about m other than (8.2), (8.3), and (8.4), i.e. whether any function m satisfying these three conditions is the marginal tax rate associated with some admissible utility function. The answer is no; $u = \log(1+x)$ is not admissible but the corresponding m (see Example 7.5) satisfies all three conditions. However, we have

Remark 8.5 *Let* m_0 *be any continuous function on the positive real numbers satisfying* (8.2) *and* (8.4). *Then for each* $K > 0$ *there is an admissible* u *whose* m *coincides with* m_0 *for* $x \leqq K$.

This says that (8.2) and (8.4) are enough to characterize marginal tax rates if we restrict ourselves to bounded net incomes. Of course, (8.3) is irrelevant when one is looking only at a bounded set of x.

To prove Remark 8.5 we can simply set

$$u(x) = \exp\left(\int_0^x \left(1 - y + \int_0^y (1 - m_0(z))^{-1} dz\right)^{-1} dy\right) - 1$$

when $x \leqq K$; when $x > K$, we adjust u smoothly so that it is bounded. Such a u satisfies all the requirements.

Remark 8.5, together with (8.2) and (8.4), provides a characterization of those functions that can appear as marginal tax rates when one restricts oneself to a bounded set of net incomes.

9 An Informal Demonstration of the Main Theorem

Though the proof of the main theorem is somewhat technical, it is possible, by using the calculus of infinitesimals, to outline the underlying ideas rather quickly. This we will now do. A completely rigorous proof will be presented in [4].

Write

$$\boldsymbol{r}_\lambda(S) = \sup\left\{\int_S \lambda u(\boldsymbol{x}) : \int_S \boldsymbol{x} = \int_S \boldsymbol{e}\right\}, \tag{9.1}$$

$$q_\lambda(S) = \begin{cases} r_\lambda(S) & \text{if} \quad \mu(S) \geqq 1/2, \\ 0 & \text{if} \quad \mu(S) < 1/2. \end{cases} \tag{9.2}$$

Intuitively, $r_\lambda(S)$ is the maximum total utility (when weighted by the $\boldsymbol{\lambda}(t)$) that S can get for itself by reallocating precisely its own endowment among its members, i.e., neither taking anything from other people nor having anything taken away from it nor destroying anything. On the other hand, $q_\lambda(S)$ is the maximum aggregate utility that S can *assure* itself.

In this demonstration we assume that $r_\lambda(S)$ is finite and that the sup is actually attained. These assumptions can be removed.

Step 1 We have

$$w_\lambda(S) = q_\lambda(S) - q_\lambda(T \setminus S). \tag{9.3}$$

An optimal pair of strategies in the threat game H_λ^s is for the majority to take from the minority everything that the minority does not destroy (*i.e., to tax at* 100 *per cent*), *and for the minority to destroy its entire endowment.*[17]

Demonstration First let S be in the majority. By taxing at 100 per cent and reallocating its own endowment, S can assure that its own payoff will be at least $r_\lambda(S)$, and that of its complement 0. Therefore it assures itself a payoff of $r_\lambda(S)$ in the game H_λ^s. On the other hand, by destroying its endowment, the minority can assure that the majority's payoff in H_λ^s will not be more than $\boldsymbol{r}_\lambda(S)$. Hence the strategies described are indeed optimal, and when $\mu(S) > 1/2$, (9.3) is proved. The argument when S is in the minority is similar, using the fact that $T \setminus S$ is then in the majority.[18]

17. Of course these threats are not carried out in the compromise that is finally reached.

18. In this heuristic treatment we ignore the case in which $\mu(S)$ is exactly 1/2. In the rigorous treatment of Aumann-Kurz [4] it is of course taken into account.

Step 2 $\phi v_\lambda = \phi q_\lambda$.

Demonstration Define the game $q_\lambda^{\#}$ dual[19] to q_λ by $q_\lambda^{\#}(S) = q_\lambda(T) - q_\lambda(T \setminus S)$. By reversing orderings in (3.3), one can easily show that $\phi q_\lambda^{\#} = \phi q_\lambda$. On the other hand, from (4.7) and (9.3) it follows that $v_\lambda = (1/2)q_\lambda + (1/2)q_\lambda^{\#}$. The result then follows from the additivity axiom for the value (4.2).

Step 3 *Suppose* $v_\lambda(T)$ *is attained at* $\boldsymbol{x}$. *Let* p *be the shadow price associated with the maximization, i.e.* $p = \boldsymbol{\lambda}(t)u_t'(\boldsymbol{x}(t))$ *when* $\boldsymbol{x}(t) > 0$. *Then*

$$(\phi q_\lambda)(dt) = \tfrac{1}{2} r_\lambda(T)\mu(dt) + \tfrac{1}{2}(\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t)) - p(\boldsymbol{x}(t) - \boldsymbol{e}(t)))\mu(dt).$$

Demonstration Let us fix attention on a particular agent; in this demonstration it will be convenient to refer to him directly as *dt*, rather than by the label *t*. Denote by *S* the set of all agents up to and including *dt* in a random order on all the agents; the value $(\phi q_\lambda)(dt)$ is the expectation of the contribution $q_\lambda(S) - q_\lambda(S \setminus dt)$ of *dt* to *S* (see (4.3)). The probability is 1/2 that *S* is in the minority, in which case *dt* contributes nothing. With probability $\mu(dt)$ (i.e. the reciprocal of the number of players), *S* is in the majority with *dt* and in the minority without him, i.e. *dt* is "pivotal." Because the ordering is random and there are many agents, *S* is almost surely an almost perfect sample of all the agents (insofar as utilities, endowments and the $\boldsymbol{\lambda}(t)$ are concerned). That is, *S* is just like *T*, but is operating at half the scale (because $\mu(S) = 1/2$). In this case, therefore, *dt*'s contribution is $r_\lambda(S) = (1/2)r_\lambda(T)$.

Finally, with probability 1/2, *S* is in the majority with and without *dt*. Again because of the random order, *S* is almost surely an almost perfect sample of the population *T* of all agents, and *dt*'s contribution $r_\lambda(S) - r_\lambda(S \setminus dt)$ is the same as if he were the last agent, i.e., the same as $r_\lambda(T) - r_\lambda(T \setminus dt)$. But now a simple computation (see for example, Aumann [3, Section 8, Step 5]) shows that the latter contribution is just

$$(\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t)) - p(\boldsymbol{x}(t) - \boldsymbol{e}(t)))\mu(dt). \tag{9.4}$$

Summing up, *dt* contributes nothing with probability 1/2, $(1/2)r_\lambda(T)$ with probability $\mu(dt)$, and (9.4) with probability 1/2. So his expected contribution $(\phi q_\lambda)(dt)$ is precisely as asserted.

Step 4 *If* $\boldsymbol{x}$ *is a value allocation, then there is a constant* c *satisfying* (5.1).

Demonstration That $\boldsymbol{x}$ is a value allocation means that $u(\boldsymbol{x})$ is a value, i.e.,

19. Compare Aumann and Shapley [5, p. 140], or Milnor and Shapley [12].

$$\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t))\mu(dt) = (\phi v_\lambda)(dt)$$

(see (4.8)). Combining this with Steps 3 and 2, we get

$$\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t)) = \tfrac{1}{2}r_\lambda(T) + \tfrac{1}{2}(\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t)) - p(\boldsymbol{x}(t) - \boldsymbol{e}(t))),$$

whence

$$\frac{1}{p}\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t)) + \boldsymbol{x}(t) = \frac{1}{p}r_\lambda(T) + \boldsymbol{e}(t). \tag{9.5}$$

From this it follows that $\boldsymbol{x}(t) > 0$, and hence $p = \boldsymbol{\lambda}(t)u_t'(\boldsymbol{x}(t))$. Inserting this into (9.5) and setting $c = r_\lambda(T)/p$, we deduce (5.1).

Step 5 If $\boldsymbol{x}$ is an allocation satisfying (5.1), *then it is a value allocation.*

Demonstration First note that $\boldsymbol{x}(t) > 0$, since $c > 0$. Hence $u_t'(\boldsymbol{x}(t))$ is defined and is positive. Set $\boldsymbol{\lambda}(t) = 1/u_t'(\boldsymbol{x}(t))$, and set $p = 1$; then (5.1) may be written in the form

$$\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t)) = c + p(\boldsymbol{e}(t) - \boldsymbol{x}(t)). \tag{9.6}$$

Integrating and recalling that $\boldsymbol{x}$ is an allocation, we obtain

$$\int \boldsymbol{\lambda}u(\boldsymbol{x}) = c.$$

Because $\boldsymbol{\lambda}(t)u_t'(\boldsymbol{x}(t)) = 1$, the integral on the left actually achieves the maximum defining $r_\lambda(T)$, i.e. $c = r_\lambda(T)$. Inserting this in (9.6) and rearranging, we find

$$\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t)) = \tfrac{1}{2}r_\lambda(T) + \tfrac{1}{2}(\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t)) - p(\boldsymbol{x}(t) - \boldsymbol{e}(t))).$$

Hence by Steps 3 and 2,

$$\boldsymbol{\lambda}(t)u_t(\boldsymbol{x}(t))\mu(dt) = (\phi v_\lambda)(dt). \tag{9.7}$$

Since $\boldsymbol{x}$ is an allocation, it follows from (9.7) and (4.8) that it is a value allocation, as was to be shown.

Step 6 There are precisely one allocation $\boldsymbol{x}$ and one real constant c satisfying (5.1), *and this constant is necessarily positive.*

Demonstration Define

$$g_t(x) = \frac{u_t(x)}{u_t'(x)} + x;$$

g_t is increasing and continuous, is defined for all nonnegative numbers, vanishes at 0, and tends to infinity as $x \to \infty$. Hence its inverse g_t^{-1} is defined and has the same properties. Moreover, $g_t(x) \geqq x$, and so

$$g_t^{-1}(y) \leqq y, \tag{9.8}$$

with strict inequality holding when $y > 0$. For each nonnegative number γ, define

$$f(\gamma) = \int_T g_t^{-1}(\gamma + \boldsymbol{e}(t))\mu(dt); \tag{9.9}$$

by (9.8) and Assumption (3.3) the integral is finite. Again using (9.8) we get

$$f(0) < \int e.$$

Next, if we go to the limit under the integration sign[20] in (9.9) and use the fact that $g_t^{-1}(x) \to \infty$ as $x \to \infty$, then we deduce that for sufficiently large γ,

$$f(\gamma) > \int \boldsymbol{e}.$$

Moreover f is strictly increasing (since the g_t^{-1} are) and continuous.[21] Hence there is one and only one γ with $f(\gamma) = \int \boldsymbol{e}$; denote this γ by c and set

$$\boldsymbol{x}(t) = g_t^{-1}(c + \boldsymbol{e}(t)).$$

By construction $\boldsymbol{x}$ is an allocation, and together with c satisfies (5.1). Conversely, if $\boldsymbol{x}$ and c satisfy (5.1), then integrating (5.1) yields $c = \int u(\boldsymbol{x})/u'(\boldsymbol{x}) > 0$, and then inverting (5.1) yields (9.10). Integrating (9.10) yields $\int \boldsymbol{e} = \int_T g_t^{-1}(c + \boldsymbol{e}(t))\mu(dt)$, and hence c and $\boldsymbol{x}$ can only be those already found. This completes Step 6.

Demonstration of the Main Theorem We have shown in Steps 4 and 5 that an allocation $\boldsymbol{x}$ is a value allocation if and only if there is a constant c satisfying (5.1). By Step 6, there are precisely one $\boldsymbol{x}$ and c satisfying (5.1); and $c > 0$. Hence there is precisely one value allocation, and it satisfies (5.1), which is what the Main Theorem asserts.

20. This may be justified by a standard theorem like Fatou's lemma or the monotone convergence theorem.

21. This follows from the fact that $|g_t^{-1}(y_2) - g_t^{-1}(y_1)| < |y_2 - y_1|$, which in turn follows from the definition of g_t.

10 Additional Discussion of the Results

In the development above we passed over issues that are important but perhaps somewhat controversial. This we did for the sake of continuity; we now wish to address ourselves to some of these issues.

(a) Aggregation

Our purpose in this paper is to try to expose the effect of political power on income redistribution. Taxation has other important aspects, such as its effect on relative prices in a multicommodity economy and aspects relating to the production of private and public goods. The reader will agree, however, that an important element of taxation is the redistributive one, even in the case in which there is no overt cash redistribution, but all revenue is used for the provision of public goods. The relatively simple model of this paper deals with the redistributive element only; other, more complex issues are left for subsequent studies.

(b) Pareto Optimality, Second Best and the Problem of Leisure

An argument may be made in connection with the fact that one of the axioms underlying the Shapley value is Pareto optimality. This means that the value allocation $\boldsymbol{x}$ is Pareto optimal, in spite of the fact that we are introducing an explicit income tax into the system. It is known that the introduction of income taxation may in theory lead individuals to take more leisure than they would take in the absence of taxation, and so may lead to outcomes that are not Pareto optimal.

Certain aspects of the nonoptimality of taxation are implicit in the fact that any agent may destroy his endowment. If we think of one component of the endowment as "labor services," then within the context of the threat game, coalitions may threaten to withhold their labor services from the market. The problem arises with regard to the final compromise embodied in the value allocation $\boldsymbol{x}$. What we are saying is that as part of the compromise, each individual promises to offer the same amount of labor services as before the bargaining process. Thus if $\boldsymbol{e}(t)$ was the endowment of agent t, this same endowment will be offered to the market at the end. The difficulty with this is that in a decentralized, democratic society an individual may choose to work as much as he wishes. He thus may accept the compromise tax function but then work less and end up with a new $\boldsymbol{e}^{\tau}(t)$ which is the taxable income arising from a tax schedule τ. The usual argument is that $\boldsymbol{e}^{\tau}(t)$ and $\boldsymbol{e}(t)$ may not be the same and the allocation based on $\boldsymbol{e}^{\tau}(t)$ may not be Pareto optimal.

Although we think there may be possibilities of reformulating our model so as to take into account these "incentive effects" of taxation, we hesitate to do so because of the rather mild empirical evidence in support of "incentive effects." Both historical and cross-section analysis indicate a negligible wage elasticity of labor supply for males and a slightly higher elasticity for females (see, for example, Hall [7] and Boskin [6]). Thus our assumption appears to be supported by the empirical evidence.

(c) Individualized Tax Schedules

Our Main Theorem indicates that the tax rate depends upon the agent's utility function as well as on his gross income; thus the solution calls for individualized tax rates. The same phenomenon arises in all welfare theoretic treatments that seek optimal income distribution. For example, the criterion of "equal marginal utility of income" entails individualized taxation. Naturally we are aware of the fact that taxes ought to be "uniform;" this question was discussed already in Section 2, in a somewhat different context (that of majority power in the threat game). As we said there, the extremely complex tax laws that one observes are in fact designed to provide some degree of individualization. Moreover, it must be remembered that we are discussing redistribution as well as taxation; though taxation may be required to be "uniform," certainly there is no way to prevent individualized redistribution.

(d) Cardinality

The cardinal nature of the results may disturb some economists, who prefer the ordinal concepts that are familiar from general equilibrium theory. The point is that in the context of a power struggle, where threats are involved, it is to be expected that intensity of preference will play a determining role. Even on a purely intuitive level, it seems clear that in a bargaining situation, the more fearful side, or the side that is more "intense"—more interested in the outcome—is also the weaker one. "Fear" and "intensity" are cardinal concepts, and it would seem that they must be explicitly taken into account in any situation involving bargaining. For this reason we would expect value theories to be necessarily cardinal; on the contrary, it is surprising that such results can be obtained without interpersonal comparisons. For a more detailed analysis, see Shapley [18].

Our cardinal results should be contrasted with those of Aumann [3], who investigated value allocations in societies using the economic mechanism of exchange only, without any political voting mechanism. One of the surprising conclusions of that study was that though cardinal utilities enter in an essential manner into the description of the model, the outcomes are independent of the choice of cardinal utilities, and depend on ordinal preferences only. The reason for this phenomenon is that in a nonatomic market situation, the marginal contribution of an agent is essentially the same almost whatever coalition he joins, so that the result of the coalitional maneuvering is a foregone conclusion and there is little or no risk. But here that is far from being the case, as there is an enormous difference between being in the majority and being in the minority. The willingness to risk being in the minority—and to have the minority make threats—is therefore of vital importance. This explains in particular why the fear of ruin is decisive, since the optimal threat of the minority is precisely to ruin its members, and only in that way can it get a fair deal from the majority.

(e) The High Tax Rates

An eyebrow-raising result of this study is that the marginal tax rate must be at least 50 per cent. This result is due to our assumption of absolute majority rule. It can be shown that if the voting rules are altered to give some weight to wealth, the tax rates tend to be smaller.[22]

In fact, the power structure of existing societies does generally give weight to wealth; i.e., wealthier individuals have more say in decision making. In a representative democracy, the extra weight given to wealth arises from the fact that in their votes, the representatives are more responsive to the stronger pressure groups: in addition to the fact that the elected representatives come from the wealthier classes, the need to use resources in running for office enables the owners of wealth to form more effective pressure groups. The result is a lower marginal tax than is indicated in this study, which might be considered the extreme case of "pure democracy."

(f) Concavity and Boundedness of the Utility Functions

Concavity is merely an assertion of general risk averseness. Boundedness is a natural assumption in the context of von Neumann–Morgenstern utilities, since unbounded utilities lead to the St. Petersburg Paradox.

(g) Interpretation of the Weights λ

The weight $\lambda(t)$ should be considered a measure of the importance of t's utility, arising endogenously out of the given power structure; indirectly,

22. In the extreme case in which decisions are reached by a majority vote of the wealth rather than the population (i.e. "money votes") there are no taxes at all; i.e. the tax rate is 0.

it is an index of t's power. Society behaves as if it were maximizing the expression $\int \boldsymbol{\lambda} u(\boldsymbol{x})$, i.e., the sum of individual utilities weighted by their relative importance. This expression resembles the social welfare function mentioned in the introduction; but the underlying idea is quite different, since the equilibrium values of $\boldsymbol{\lambda}$ change if we change the initial distribution of wealth or if we change the utilities. But apart from that we eschew all the paternalistic, ethical connotations of the phrase "social welfare," and so prefer here to call $\int \boldsymbol{\lambda} u(\boldsymbol{x})$ a *social power function*. Society maximizes total weighted utility, where the more powerful individuals are considered more important—i.e., are given more weight. Of course we are not advocating this (or any other) procedure; but most observers will agree that in fact, society takes much less account of "equity" or "fairness" than of power.

This may at first sound cynical, but when one examines its implications more carefully it turns out to be quite the opposite: a reaffirmation of the importance of democracy. Since social decision making is a function of power, it follows that to achieve equity in the outcome, one must build equity into the political institutions. A system that concentrates power in the hands of the few, even with the most idealistic intentions, must in the end benefit the few. It has long been known that income distributions in countries with rightist totalitarian regimes are more skewed than in the western democracies; but as Wiles [21] has shown,[23] this is to some extent true also for leftist totalitarian regimes (once one removes the cosmetic layers under which official figures are buried). So it is not ideology that is decisive, but the power structure; and if we want a more equitable society, we must develop institutions that spread the political power as thinly and evenly as possible.

References

1. Arrow, K. J.: *Aspects of the Theory of Risk-Bearing*. Helsinki: Yrjö Jahnssonin Säätiö, 1965.

2. Arrow, K. J., and M. Kurz: *Public Investment, the Rate of Return, and Optimal Fiscal Policy*. Baltimore: Johns Hopkins Press, 1970.

3. Aumann, R. J.: "Values of Markets with a Continuum of Traders," *Econometrica*, 43 (1975), 611–646 [Chapter 51].

4. Aumann, R. J., and M. Kurz: "Power and Taxes in a Multi-Commodity Economy," *Israel Journal of Mathematics*, 27 (1977), 185–234 [Chapter 53].

5. Aumann, R. J., and L. S. Shapley: *Values of Non-Atomic Games*. Princeton: Princeton University Press, 1974.

23. We are indebted to E. Sheshinski for this reference.

6. Boskin, M. J.: "The Economics of Labor Supply" in *Income Maintenance and Labor Supply*, ed. by G. G. Cain and H. W. Watts. New York: Academic Press, 1973.

7. Hall, R. E.: "Wages, Income, and Hours of Work in the U.S. Labor Force" in *Income Maintenance and Labor Supply*, ed. by G. G. Cain and H. W. Watts. New York: Academic Press, 1973.

8. Harsanyi, J. C.: "Approaches to the Bargaining Problem Before and After the Theory of Games: A Critical Discussion of Zeuthen's, Hicks's and Nash's Theories," *Econometrica*, 24 (1956), 144–157.

9. ———: "A Bargaining Model for the Cooperative *n*-Person Game" in *Contributions to the Theory of Games IV*, ed. by R. D. Luce and A. W. Tucker. Princeton: Princeton University Press, 1959.

10. Hildenbrand, W.: *Core and Equilibria of a Large Economy*. Princeton: Princeton University Press, 1974.

11. Kannai, Y.: "Values of Games with a Continuum of Players," *Israel Journal of Mathematics*, 4 (1966), 54–58.

12. Milnor, J. W., and L. S. Shapley: "Values of Large Games: Oceanic Games," *Mathematics of Operations Research*, 3 (1978), 290–307.

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

14. Nash, J. F.: "The Bargaining Problem," *Econometrica*, 18 (1950), 155–162.

15. ———: "Two Person Cooperative Games," *Econometrica*, 21 (1953), 128–140.

16. Pratt, J. W.: "Risk Aversion in the Small and in the Large," *Econometrica*, 32 (1964), 122–136.

17. Shapley L. S.: "A Value for *n-Person Games" in Contributions to the Theory of Games II*, ed. by H. W. Kuhn and A. W. Tucker. Princeton University Press, 1953.

18. ———: "Utility Comparisons and the Theory of Games," in *La Décision*. Paris: Editions du Centre National de la Recherche Scientifique, 1969.

19. Sheshinski, E.: "The Optimal Linear Income-Tax," *Review of Economic Studies*, 39 (1972), 297–302.

20. Von Neumann, J., and O. Morgenstern: *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.

21. Wiles, P.: *Distribution of Income: East and West*. Amsterdam: North-Holland, 1974.