COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 544

RAWLSIAN JUSTICE AS THE CORE OF A GAME

John Roemer and Roger Howe

February 28, 1980

RAWLSIAN JUSTICE AS THE CORE OF A GAME*

by

John E. Roemer and Roger E. Howe

Abstract: It is suggested that the ethical notion of social contract can be formally modeled using the well studied concept of the core of a game. This provides a mathematical technique for studying social contracts and theories of justice. The idea is applied to Rawlsian justice here.

1. Justice, Social Contracts, and Games

A natural way of formally describing the idea of social contract is as the core of a game. Consider the members of society as the participants in a game with many players. Every coalition of people could, on its own, make available to its members a certain amount of income (or goods) if it withdrew from society and operated as a mini-society on its own. A game (in coalitional form) may be defined by specifying what the income (or pay-off) would be to every conceivable coalition, if it were to withdraw from society. Consider, now, an allocation of income, an income distribution, for all of society. If there is no coalition which can better the lot of all its members by exercizing its withdrawal option, then the given income distribution is said to be in the core of the game. The core is the set of all income distributions which cannot be blocked by any coalition by withdrawal as described.

---

*The authors wish to acknowledge the useful comments of Michael Woodford. Roemer's work is partially supported by NSF Grant #SPI-79-14889, and Howe's work by NSF Grant #MCS-79-05018.

We say the core models the idea of social contract because it captures the conflicting notions of cooperation and necessity which are at the heart of the contractarian notion. An allocation in the core is one which all members agree to abide by, not because it is some sense marvelous for every individual, but because there is no conceivable way that any group can organize to gain a superior position for its members, given the rules of organization to which society's members have agreed. Clearly what differentiates one social contract from another are the rules of coalition organization that society specifies which embody its ethics. By rules of organization we mean this: Society must adopt a rule for assigning any coalition its "dowry," should it decide to withdraw; these pay-offs to coalitions define the game. For instance, society might decide that each coalition, if it withdraws, can take with it just the original endowments of its members. Or, alternatively, society might decide that any withdrawing coalition can take with it its per capita share of total social endowment. These two withdrawal rules define two different games, with two different social contracts; there are, of course, many more possibilities. The rule by which society specifies the hypothetical dowry of any coalition embodies society's ethical notion of what coalitions are entitled to, should they decide to opt out of society. In the first rule proposed above, the ethical premise is that each person has a right to his own endowments; in the second rule, each person is supposed to have an equal claim on society's total pool of endowments. The core of the game is then the collection of stable social contracts, as at a core allocation no coalition will have any incentive to withdraw from society, given the ethical presumption concerning its entitlement if it decided to go it alone.

The rule which society adopts to define the hypothetical dowries or pay-offs to its coalitions embodies what has been called above its ethical presumption, or its concept of justice. That is, a concept of distributional justice entails specifying what entitlements individuals, and more generally coalitions of individuals, should have, were they to complain about their rewards in actual society, and opt out. In this language, a concept of justice is described by specifying the game (in the sense of specifying the dowries of all coalitions); an allocation in the core of that game is then a <u>just allocation</u>, as it cannot be improved upon by any coalition under its hypothetical alternative of exercising the ethical right to its dowry. This formulation of justice explicitly exhibits the <u>contractarian nature</u> of a theory of justice: for we have just concluded that a just allocation is in the core of a game, and is therefore a social contract.

These observations invite the following procedure. Given a theory of justice, one may ask: To what game does this theory of justice correspond? What is the game whose core consists of precisely the "just allocations" of the theory in question?

In this paper, we apply this method to Rawlsian justice. When Rawls maintains that rational individuals would agree, in a certain pristine state, to a particular income distribution (the maximin one), he is suggesting that there is a social contract which gives rise to the maximin distribution. If we think of individuals as taking part in a lottery for natural endowments, behind the veil of ignorance, the outcome of which will determine their potential for generating income ("the good"), the game structure of the problem becomes evident. We will allow an individual or a coalition of individuals to express its dissatisfaction with a particular income distribution by hypothetically withdrawing from society, and testing whether under the rules of the game it can improve the lot of its members. We seek rules of the game which will generate as the core precisely the maximin income distribution(s).

This approach allows us to evaluate the Rawlsian theory of justice in a somewhat detached way, since the ethical content of the theory must be embodied in the criteria for withdrawal which define the game. In fact, we shall see the Rawlsian criteria are only one possible set of such criteria, and that other apparently reasonable withdrawal criteria imply other social contracts, other theories of justice, in the Rawlsian environment.

In this paper, we model only Rawlsian justice. Clearly, the methodology is quite general, and one can ask how other notions of justice would appear in explicit form, as it were, as the rules of a game. One of the authors has carried out elsewhere this procedure for a number of theories of justice (Roemer (1980)), with an emphasis on the Marxian theory of exploitation.

2. A Model of the Rawlsian Problem

To model the Rawlsian problem, it will be convenient to view individuals in society as points in a continuum, say the interval $[0,1]$, indexed by the variable t. There is an index of the endowment which an individual might receive, called α, which ranges over some interval $[a,b]$ and is assigned to agents under some probability distribution f. Behind the veil of ignorance, individuals do not know how much α they will receive, but they know there is such a probability distribution f.

There is a production function which allows individuals to convert some level of endowment α into income Y, or "the good." Income produceable is an increasing function of α . We shall be concerned with redistributing these incomes which are produced initially once people get their endowment assignments and shall model this redistribution as a tax on income. The incentive problem takes this form: if the tax rate on an individual's income rises, he produces less pre-tax income. Thus we specify a <u>production-incentive function</u> $g(\alpha,\tau)$, which specifies how much income an individual of endowment α will produce if he is taxed at rate τ . His after-tax income will be $(1-\tau)g(\alpha,\tau)$. The function g is the same for all members of society. Behind the veil of ignorance, everyone knows there is such a function g (though no one needs to know what the particular function is). The incentive effect takes this form:

$$\frac{\partial g}{\partial \tau}(\alpha,\tau) < 0 \text{ for } \tau > 0 \text{ and } \frac{\partial g}{\partial \tau}(\alpha,\tau) > 0 \text{ for } \tau < 0 \qquad (*)$$

A <u>feasible tax scheme</u> $\tau(t)$ is a system of redistribution which precisely balances: the tax revenues should equal the expenditures (to the subsidized). (If $\tau(t) < 0$, then t receives a <u>subsidy</u> in the amount $\tau(t)g(\alpha(t), \tau(t))$. It is still true for subsidized agents t that their after-tax income is $(1-\tau)g(\alpha,\tau)$.)

<u>Definition</u>. A function $\tau(t)$ on $[0,1]$ is a <u>feasible tax scheme</u>, if

$$\int_0^1 \tau(t)g(\alpha(t), \tau(t))d\alpha(t) = 0 .$$

Each feasible tax scheme gives rise to a post-tax income distribution. We formulate now the definition of the maximin distribution(s). Given a feasible tax scheme $\tau(t)$, an income distribution is produced with a certain smallest after-tax income. The problem is to find that distribution which maximizes this minimum income. We call this maximin income Y_{min} , defined by:

$$Y_{min} = \max_{\substack{\tau(t) \\ \text{feasible}}} \min_{t} (1-\tau(t))g(\alpha(t), \tau(t)) .$$

We shall study this model in detail in Section 6 below, where we show a unique maximin tax scheme exists, and where some of its properties are elucidated. Note that the income distribution problem is fully specified by the functional knowledge $\{f,g\}$. These functions determine the tax schemes and hence the feasible income distributions.

3. Rawlsian Justice as the Core of a Game

The veil of ignorance is the key notion in Rawlsian justice, and we formulate a game involving it as follows. Individuals behind the veil of ignorance wish to choose a tax scheme $\tau(t)$ (or choose rules for choosing a tax scheme if they do not know the functions f and g) with the following property: that after the lottery for α is taken, and after post-tax incomes are realized according to the chosen tax scheme, no coalition of agents would be able to guarantee an improvement in after-tax income for all its members by withdrawing again behind the veil of ignorance and trying another draw on the lottery f . The criteria for withdrawal which define the game are: a coalition blocks an allocation if it can guarantee an allocation of income for its members which improves all its agents' incomes, after withdrawing again behind the veil of ignorance and taking the outcome of the lottery performed for the coalition from the same distribution of skills f . We call this kind of block an ERA-block, and the game is the ERA-game. (ERA stands for extremely risk averse. The motivation shall be clear later.)

Several technical points need be mentioned. First, the only coalitions which are taken into account are those of positive measure. The desires of a null coalition (of zero measure) are considered irrelevant. This is standard procedure in the treatment of games with a continuum of agents, and indeed it captures Rawls's intention to consider only social groups of significant size. Since we only consider coalitions of a positive measure, we will wish to assure ourselves that the agents receiving Y_{min} in a maximin distribution are in fact a coalition of positive measure: for if they were not, the focus of Rawls's theory, those agents, would be irrelevant from the model's point of view. We shall prove in a later section that:

Lemma. Under reasonable conditions, the coalition of agents receiving after-tax income Y_{min} in a maximin distribution, has positive measure.

The second technical point is that we insist a blocking coalition be able to improve the income of almost every (a.e.) member. Sometimes in economics a coalition is allowed to block an allocation if it can improve the income of some members, and not reduce the income of any member. However, one should note that this weak form of blocking is only justifiable, when it is used, if it can be transformed into strong blocking, by taking a little of the increment in income of the strictly better off members, and redistributing it to all the other members. In our model, this redistribution is illegitimate, since taking a little income from one member will affect his incentive, and change the amount he produces. Thus, the importance of incentives in the Rawlsian problem nullifies the equivalence of weak and strong blocking. The proper notion of blocking must be strong blocking, where a.e. member can improve his income.

We now prove that the ERA-game is equivalent to the Rawlsian theory of justice.

Definition. A distribution is maximin if it has at most a null set of agents receiving incomes less than Y_{min} .

Theorem. The core of the ERA-game consists precisely of the maximin distributions.

Proof. We show first that any maximin distribution is in the core. Let $Y(t) = (1-\tau(t))g(\alpha(t), \tau(t))$ be such a distribution and consider the coalition W of agents who receive income Y_{min} or less:

$$W = \{t | Y(t) \leqq Y_{min}\} .$$

From the Lemma it follows W is a coalition of positive measure. We show W cannot block the distribution $Y(t)$. If W tries the lottery again, the production possibilities it has open to it are the same as society as a whole enjoys--that is, the same production-incentive function $g(\alpha,\tau)$ applies to W as to the universe $[0,1]$. (This assumption may be thought of as a form of constant return to scale: W , being itself a continuum, is simply a scaled-down version of society, and has therefore the same function g available to it, as well as f .) After the lottery is taken for W , the coalition can assign itself various income distributions. If it assigns itself a maximin distribution, then there will be a subcoalition $W^* \subseteq W$ of positive measure which receives precisely Y_{min} , by the Lemma. Since W^* is non-null, almost all its agents were receiving Y_{min} in society's original lottery, and so they were not made better off by the withdrawal. If W chooses for itself an allocation which is not maximin, then some non-null subcoalition of

W receives income strictly less than Y_{min}, and hence those agents are strictly worse off than they were in the original allocation $Y(t)$. Hence the coalition W cannot ERA-block $Y(t)$, as there is no allocation of income in W which improves thelot of almost every member.

But if W cannot block $Y(t)$, then no non-null coalition of $[0,1]$, can block $Y(t)$, since any other candidate for a blocking coalition must contain a non-null set of agents who receive more than Y_{min} in the original distribution $Y(t)$. A fortiori, such a coalition cannot guarantee an improvement for its members by going back behind the veil of ignorance.

Conversely, let $\hat{Y}(t)$ be any income distribution in the ERA-core, and suppose $\hat{Y}(t)$ is not maximin. Then consider the coalition $\hat{W}$ of agents who receive less than Y_{min} under $\hat{Y}(t)$. By supposition, $\hat{W}$ is a non-null coalition. Assign to $\hat{W}$ a maximin distribution. Then every member of $\hat{W}$ is strictly better off, and $\hat{Y}(t)$ is ERA-blocked.

Q.E.D.

It should be pointed out that the Theorem can be proved without resorting to the Lemma. We present the above proof, however, since the Lemma is true, and since the other proof is somewhat less intuitive.[1]

4. A Risk Neutral Game

To gain some perspective into the Rawls game, we propose as a foil a risk neutral game. In the RN-game, the agents wish to maximize expected income. Behind the veil of ignorance an income distribution scheme $Y(t)$ is proposed. A coalition of agents will withdraw from the lottery if, under probability distribution f and production-incentive function g, members can assign themselves a post-lottery distribution in which every member can expect to do better than he can expect to do in $Y(t)$. This is a risk neutral game, since agents are concerned only with maximizing expected income.

Note that the expectation of income is maximized under that distribution scheme which maximizes society's total income. But the unique distribution which maximizes total income is the one with no redistribution $(\tau(t) \equiv 0)$, since any taxation reduces production, by the incentive assumption (*). Consequently, the RN-core consists precisely of the distributions which are "a.e. no-taxation"--that is, mathematically equivalent to the no redistribution scheme. This is, to say the least, a very different social contract from the ERA game.

It can now easily be seen how the ERA game is risk averse. A coalition of agents in the ERA game is only willing to block an income distribution if, with probability one, its members can be made better off. The coalition members are not willing to take any risks of a deteriorated position. In the RN game, agents will block if their mathematical expectation is to improve their position.

There is an interesting consequence of this difference between the two games. In the ERA game, agents will have no desire to re-contract, after the lottery is taken and the maximin distribution is assigned. That is, no coalition could assure an improved position for almost all its members. In the RN game, however, there will be some losers after the lottery is drawn and the no-tax scheme assigned. The coalition of agents who receive less than the mean income would like to re-contract, as they could expect to improve their position with a new draw. This distinction points out the essence of risk aversion: the extremely risk averse agents take no risks, so they will have no regrets; some agents in the RN game who take risks will be disappointed. There will be coalitions willing to organize a new lottery under the RN rules, but no such coalition will exist under the ERA rules. This points out why people must pass judgment on the acceptability of an income distribution proposal before the lottery has been drawn. If there is any risk in the agents' evaluation of the distribution, there will by definition be some who will wish to re-contract after the lottery. The ERA rules are the only ones in which no coalition will wish to re-contract after lottery, and this is precisely because no risks are taken under ERA rules. However, the once-and-for-all nature of the Rawlsian lottery does not imply that rational individuals must be extremely risk averse.

Total income in the maximin distribution is less than in the no-tax distribution. The difference between society's income in the two cases can be viewed as the premium which members of the ERA society pay to insure themselves against the luck of the draw.

5. A Continuum of Justice Games

We can define the ERA game in a different way. Suppose agents wish to choose the distribution which maximizes their expected income, under the constraint that the distribution assigns every agent an income of at least Y_{min}. This specification gives rise to the maximin distribution, as will be shown in the next section.

We have already seen that the core of the RN game is the distribution which maximizes expected income, subject to no constraint on the lower bound of income received by individuals.

There is a natural intermediate case. Let Y_0 be a value of income which is less than Y_{min} but more than the worst-off person receives in the no-tax distribution:

$$Y_{min} > Y_0 > g(a,0) \ .$$

Suppose agents wish to choose a distribution which maximizes expected income subject to the constraint that they receive at least Y_0 . It is shown in Section 6 that there is a unique distribution fulfilling this requirement. These distributions reflect an intermediate type of risk aversion between the ERA and RN cases.

We can thus imagine a continuum of justice games; the ERA-game with its core of the maximin distribution is at one extreme, and the RN-game with its core consisting of the no-redistribution distribution is at the other extreme. Each of the games in this continuum can be thought of as one whose core describes the social contract for a society with some degree of risk aversion, going from totally risk averse to risk neutral.

6. A Mathematical Formulation

In this section we indicate the nature of the solution to the redistribution problem of Section 2. Let us review the formulation. We suppose a person's talents specified by some parameter α ranging between 0 and 1, and from no one consider an agent to be identified by his talent, α . The range $[0,1]$ is purely conventional. As long as one requires only measurability, not continuity, of the functions one is dealing with, the interval $[0,1]$ may be substituted for any more complicated,conceptually more appropriate parameter space.[2] We further suppose that there is a function $g(\alpha,\tau)$ that describes the total revenue that will be produced by someone with talent α when he is taxed at a rate τ . The amount of tax revenue produced is then $\tau g(\alpha,\tau)$, and the person's net income is $(1-\tau)g(\alpha,\tau)$. Of course τ can be negative, so that there can be a net transfer of income to someone. The function g may be presumed measurable in α and continuous in τ.

In adopting this description of income we are assuming that a given person's behavior depends only on how he is treated, not on how others are treated. What a person produces depends only on his tax rate, not on the tax structure as a whole. There are no invidious comparisons made, no taxation externalities.

We will also assume that $g(\alpha,\tau)$ is differentiable in τ, and that

$$\frac{d}{d\tau}(1-\tau)g(\alpha,\tau) < 0 \,. \qquad (**)$$

That is, each person's after tax income decreases as the rate of taxation increases.[3] This assumption puts limits on a person's response to taxation. It says the more one is taxed, the less one's income, and, reciprocally, the less one is taxed, or the more negatively one is taxed, the higher one's income. Various bizarre, though conceivable, types of behavior, are thereby precluded. On the one hand, we disallow the extreme reaction to taxation whereby one is driven to work so much harder than before taxes that one's after tax income actually increases, on the other hand, we disallow extreme laziness or failure of spirit on the part of the beneficiaries of taxation, so that they never work so much less after receiving tax aid that their income actually goes down. Clearly neither kind of behavior is expected.

We will also assume there is some floor on earnings, regardless of talent and rate of taxation. That is, we assume $g(\alpha,\tau) \geq \delta > 0$ for all α and τ.

The task is to find a taxation scheme $\tau(\alpha)$ which maximizes the minimum income of society. That is, we want to specify a feasible tax scheme $\tau(\alpha)$:

$$\int_0^1 \tau(\alpha) g(\alpha, \tau(\alpha)) d\alpha = 0$$

so that, subject to the above condition, the quantity

$$\min_{\alpha} (1 - \tau(\alpha)) g(\alpha, \tau(\alpha))$$

is as large as possible.

To see what the solution to this problem looks like, consider a proposed optimal tax scheme $\tau(\alpha)$. We can divide society into two groups:

$$C^+ = \{\alpha : \tau(\alpha) \geq 0\}$$

$$C^- = \{\alpha : \tau(\alpha) < 0\} .$$

Basically, what our taxation scheme does is collect money from people in C^+ , and distribute it among the people in C^- . The total amount of money collected is

$$\int_{C^+} \tau(\alpha) g(\alpha, \tau(\alpha)) d\alpha .$$

The total disbursed is

$$-\int_{C^-} \tau(\alpha) g(\alpha, \tau(\alpha)) d\alpha .$$

Clearly, if there is to be any redistribution, both C^+ and C^- must be of positive measure.

We see that, having chosen C^+ and C^-, our problem decomposes into two separate problems. In C^+ we are concerned only with raising money, in C^- only with giving it away. Let us consider this second problem first, since it is by giving money to the poorest that we will achieve our goal.

Proposition. The after tax income $(1-\tau(\alpha))g(\alpha, \tau(\alpha))$ of everyone in C^- is the same if τ is optimal.

Proof. Suppose not. Then some will be better off than others. But then we can distribute somewhat less to these people. (By definition everyone in C^- receives some tax aid.) This will free up some funds which can now be distributed to among those who are worst off. Our assumptions on g guarantee this redistribution will actually result in improving the lot of those who receive the redistributed funds. Since these were the poorest, under the new scheme, the poorest are richer than they were under the previous scheme, contradicting its optimality. We have stated the argument roughly, but under our assumptions it can easily be made precise.

Hence under an optimal tax scheme, all who are taxed negatively are raised to exactly the same level of income. Call this income level I_0. In particular, since C^- is a set of positive measure, we have also proved the Lemma of Section 3.

Now consider C^+. Here we must decide how to tax individuals in order to raise the money for distribution in C^-. Suppose we set individual α's tax rate at $\tau(\alpha)$. Why would we not increase $\tau(\alpha)$? There can be only two reasons: i) either we have already reduced α's income to the base level I_0, so further taxation of α would make our solution sub-optimal; or ii) further taxation would so destroy α's initiative that it would produce less tax revenue rather than more. Therefore $\tau(\alpha)$ is set according to the following criterion:

Either (i) $(1-\tau(\alpha))g(\alpha, \tau(\alpha)) = I_0$,

or (ii) $(1-\tau(\alpha))g(\alpha, \tau(\alpha)) > I_0$, and

$$\tau(\alpha)g(\alpha, \tau(\alpha)) = \max_{\tau} \tau g(\alpha,\tau) .$$

The criterion (ii) for maximizing tax revenue may be expressed in terms of the usual condition on derivatives

$$\tau\frac{\partial g(\alpha,\tau)}{\partial\tau} + g(\alpha,\tau) = 0 .$$

This equation can typically be expected to define a unique value for $\tau(\alpha)$, though it may not always do so. The first condition (i) will, because of our assumption on g , always define a unique value for τ .

Having described the solutions in C^- and C^+ , we can describe how C^- and C^+ are determined. Suppose we decide on a certain minimum income level I . Then everyone whose "natural" income $g(\alpha,0)$ is below I will have to be taxed negatively to achieve income level I . The rest may be taxed positively. Thus we set

$$C^-(I) = \{\alpha : g(\alpha,0) < I\}$$

$$C^+(I) = \{\alpha : g(\alpha,0) \geq I\} .$$

For each α in $C^-(I)$, we can determine the level of negative taxation $\tau(\alpha)$ necessary to raise α to income level I by the equation

$$(1-\tau(\alpha))g(\alpha, \tau(\alpha)) = I , \quad \alpha \in C^-(I) .$$

Our assumptions on g guarantee this equation has a unique solution. The total amount needed for distribution to $C^-(I)$ is then

$$-\int_{C^-(I)} \tau(\alpha)g(\alpha, \tau(\alpha))d\alpha \ .$$

To raise this, we apply the taxation scheme in $C^+(I)$ as described above. This will produce revenue

$$\int_{C^+(I)} \tau(\alpha)g(\alpha, \tau(\alpha))d\alpha \ .$$

If I is set low, the set $C^-(I)$ will be small, and relatively little revenue will be needed to raise the level of people in $C^-(I)$ to I. On the other hand, $C^+(I)$ will be large and the taxation scheme will produce a large amount of revenue for redistribution. For very large I, the situation will be the reverse--large amounts of revenue will be required, and little will be forthcoming. Evidently, then, for some intermediate value I_0, the required and available revenue will match: it is for this equilibrium I_0 that we will have

$$C^- = C^-(I_0) \ , \quad C^+ = C^+(I_0) \ ,$$

and this I_0 will be the maximin solution to the distribution problem.

In sum, then, the general shape of the income profile will be as follows.

i) A large group of people will receive the base income I_0. This will include all people whose "natural" income would be less than I_0, together with those who have been taxed until their after tax income is reduced to I_0.

ii) There may be some people whose income is larger than I_0 . These are people who have been "milked dry" for tax purposes, in the sense that increasing their tax rate would not only make them poorer, but would actually produce less tax revenue for redistribution.

The above solution suggests a family of related redistribution schemes which would not maximize the minimum income, but would guarantee some minimum level of income, and optimize total income subject to that guarantee. Thus we could set some desired minimal income level $I < I_0$ and form $C^-(I)$ and $C^+(I)$ as above. We would increase the incomes of people in $C^-(I)$ just as before. But now our all-out taxation scheme in $C^+(I)$ will produce more revenue than we need so we are free to institute an alternate, more lenient tax law. The most obvious rule to apply is to tax people so that their total income

$$\int g(\tau, \tau(\alpha))d\alpha$$

is maximal, subject to everyone having at least I in after-tax income, and subject to raising the required amount. In particular, by maximizing total income, we assure that the income distribution is Pareto optimal, subject to the requirement that no one receive less than the floor I .

How is the optimal tax rate to be determined? We will give a differential condition which may typically be expected to provide a well-defined answer. Suppose we contemplate an alteration in our tax structure, from $\tau(\alpha)$ to $\tau(\alpha) + \varepsilon(\alpha)$. Then the change in tax revenue is

$$\int_{C^+} \tau(\alpha)g(\alpha, \tau(\alpha))d\alpha - \int_{C^+} (\tau(\alpha) + \varepsilon(\alpha))g(\alpha, \tau(\alpha) + \varepsilon(\alpha))d\alpha$$

$$\sim \int_{C^+} \varepsilon(\alpha)\left[g(\alpha, \tau(\alpha)) + \tau(\alpha)\frac{\partial g}{\partial \tau}(\alpha, \tau(\alpha))\right]d\alpha \quad . \qquad (1)$$

On the other hand, the change in total income is

$$\int_{C^+} g(\alpha,\ \tau(\alpha))d\alpha - \int_{C^+} g(\alpha,\ \tau(\alpha)+\varepsilon(\alpha))d\alpha$$

$$\sim \int_{C^+} -\varepsilon(\alpha)\frac{\partial g}{\partial \tau}(\alpha,\ \tau(\alpha))d\alpha\ . \tag{2}$$

Optimality requires that whenever ε is such as to make the first integral (1) zero, the second (2), should be ≥ 0 . But since ε can be replaced with $-\varepsilon$, this just says the second integral should be zero also. This means the coefficients of ε in the two integrals must be linear multiples of one another. Thus there is some constant c such that

$$c\frac{\partial g}{\partial \tau}(\alpha,\tau) = g(\alpha,\tau) + \tau\frac{\partial g}{\partial \tau}(\alpha,\tau) \tag{3}$$

which may be rewritten:

$$g(\alpha,\tau) = (c-\tau)\frac{\partial g}{\partial \tau}(\alpha,\tau)\ . \tag{4}$$

An economic interpretation of the optimality condition for choice of τ is clear from (3): the ratio of the marginal change in tax revenue to the marginal change in total revenue should be the same, c , for all people. The usual microeconomic argument verifies this is the proper marginal condition for optimality of τ . Evidently, the constant c can be adjusted to produce the necessary revenue, consistent with the requirement that after tax income be at least I .

Thus we have a continuum of tax schemes, to produce any minimum income up to I_0 , and to maximize total income subject to the specified minimum. When we let the required floor be I_0 , we produce the Rawlsian maximin solution; when there is no floor, we produce the risk-neutral income distribution discussed in Section 4. For values of I less than I_0, the income distribution corresponds to the core of a game where individuals have some intermediate degree of risk aversion between risk neutrality (RN) and extreme risk aversion (ERA).

7. Conclusion

Our main result is methodological: we hope that exhibiting the connection between theories of justice, social contracts, and cores of games will prove fruitful in allowing comparisons of theories of justice. The game formulation of a theory of justice expresses the theory in a canonical way, and different theories can therefore be compared dispassionately by examining the rules of their respective games. For instance, we have observed that Rawlsian justice is predicated on the assumption that society's members are extremely risk averse--or act "as if" they are. It appears, for another instance, that the maximin theory of justice differs from the "individualistic" theory of justice (where there is no taxation and everyone keeps the income he generates from his endowments) not so much in the social ethics concerning what people are entitled to under the two theories but rather in the different assumptions made in regard to the degree of risk aversion of the participants. (That is, the risk-neutral game (RN) leads to the no-tax income distribution.) But this last conclusion is better recast in this way: in the Rawlsian environment, the ethical position which society adopts concerning the entitlements of its members is implied by its conception of risk which its members should or do have. This isomorphism between theories of justice ranging from "maximin" to "individualistic," on the one hand, and the degrees of participants' risk aversion, on the other, is a perhaps surprising outcome of the analysis, and shows we cannot treat postulates concerning risk aversion of agents as ethically neutral.

FOOTNOTES

[1]A brief proof without the Lemma: Suppose in the maximin distribution $Y(t)$ the coalition of worst-off agents $W = \{t | Y(t) \leqq Y_{min}\}$ were measure zero. Then any candidate coalition C for a blocking coalition, which must be non-null, has almost all its agents better off than Y_{min}. In particular, there is a sub-coalition C^* of C, of positive measure, whose agents are wealthier than $Y_{min} + \varepsilon$. By definition of Y_{min}, in any allocation which C assigns itself, there will be a set of positive measure receiving incomes less than $Y_{min} + \varepsilon$. Thus there is a positive probability that any agent in C^* could be worse off than he is originally. Hence no distribution over C can guarantee (assure with probability one) an improvement for almost all its members.

[2]Note we have eliminated the parameter t, used to index agents in Section 2, and are now indexing agents by their talent level α. The distribution f of talents across agents is now implicit in the parameterization of agents by their talent.

[3]Note that assumption (**), which is needed to derive the form of the maximin distribution studied below, differs from the incentive assumption (*). In particular, (*) implies (**) for $\tau>0$; but for $\tau<0$, neither (*) nor (**) implies the other. Nevertheless, both assumptions hold for a typical conception of rational, self-interested agents. We shall not discuss what underlying utility function for income and leisure would generate these reactions to taxation.

REFERENCES

J. Rawls, A Theory of Justice, Harvard University Press, 1971.

J. E. Roemer, "Inequality, Exploitation, Justice and Socialism: A Theoretical-Historical Approach," Cowles Foundation Discussion Paper 545, Yale University, February 1980.